Access DB# _/_

SEARCH REQUEST FORM

Scientific and Technical Information Center

FEB 21

Requester's Full Name: _John Ulm_ Examiner # : _6750?_ Date: _2-21-02_
Art Unit: _1646_ Phone Number 30 _3-408_ Serial Number: _09/509 132_
Mail Box and Bldg/Room Location: _loc C1,CM1_ Results Format Preferred (circle): **PAPER** DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.
****

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

_____

Earliest Priority Filing Date: _____

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Please search SEQ ID no:1
of 09/509482.

Point of Contact:
Beverly Shears
Technical Info. Specialist
CM1 1E05 Tel: 308-4994

mllc-/

****

STAFF USE ONLY	Type of Search	Vendors and cost where applicable
Searcher: _Beverly e4994_	NA Sequence (#)_____	STN _____
Searcher Phone #: _____	AA Sequence (#)_____	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: _____	Bibliographic _____	Dr.Link _____
Date Completed: _02-25-02_	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: _3_	Fulltext _____	Sequence Systems _____
Clerical Prep Time: _____	Patent Family _____	WWW/Internet _____
Online Time: _20_	Other _____	Other (specify)_CGN_

PTO-1590 (8-01)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2002, 07:37:10 ; Search time 1444.76 Seconds
(without alignments)
1096.188 Million cell updates/sec

Title: us-09-509-482-1
Perfect score: 96
Sequence: 1 gtttcctctctgtcgggg...........cagaactgctggagtggagg 96

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 824859755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : GenEmbl:*
     1: gb_ba:*
     2: gb_htg:*
     3: gb_in:*
     4: gb_om:*
     5: gb_ov:*
     6: gb_pat:*
     7: gb_ph:*
     8: gb_pl:*
     9: gb_pr:*
    10: gb_ro:*
    11: gb_sts:*
    12: gb_sy:*
    13: gb_un:*
    14: gb_vi:*
    15: em_ba:*
    16: em_fun:*
    17: em_hum:*
    18: em_in:*
    19: em_om:*
    20: em_or:*
    21: em_ov:*
    22: em_pat:*
    23: em_ph:*
    24: em_pl:*
    25: em_ro:*
    26: em_sts:*
    27: em_sy:*
    28: em_un:*
    29: em_vi:*
    30: em_htgo_hum:*
    31: em_htgo_inv:*
    32: em_htgo_rod:*
    33: em_htg_hum:*
    34: em_htg_inv:*
    35: em_htg_rod:*
    36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
1		96	100.0	96	9	HOMOVDR1	AF080454 Homo sapi
2		30	31.2	1119	5	AF109063	AF109063 Homo sapi
3		30	31.2	1119	5	AF110424	AF110424 Camarhync
4		30	31.2	95375	9	AL023322	AL023322 Human DNA
5		29.8	31.0	111937	9	HS228A9	HS228A9 Homo sapi
6		29.4	30.6	110000	2	AC006349	AC006349 Homo sapi
7		29.4	30.6	212827	9	AC092439	AC092439 Homo sapi
8	c	29.2	30.4	854	2	AC026090	AC026090 Homo sapi
9		29.2	30.4	128361	9	AC009412	AC009412 Homo sapi
10	c	29.2	30.4	172027	2	AC080964	AC080964 Giardia i
11	c	29.2	30.4	150771	2	AC004877	AC004877 Homo sapi
12	c	29.2	30.4	154998	9	AC040944	AC040944 Homo sapi
13	c	29.2	30.4	163447	9	AL596186	AL596186 Homo sapi
14	c	29.2	30.4	174476	9	AL354977	AL354977 Human DNA
15	c	29.2	30.4	198606	2	AL353577	AL353577 Homo sapi
16		29	30.2	72081	2	AC092000	AC092000 Homo sapi
17	c	29	30.2	72084	2	AC090409	AC090409 Homo sapi
18	c	29	30.2	163106	2	AC016167	AC016167 Homo sapi
19	c	29	30.2	142126	2	AP003375	AP003375 Homo sapi
20	c	29	30.2	170676	9	AC018970	AC018970 Homo sapi
21	c	29	30.2	175565	2	AP002769	AP002769 Homo sapi
22	c	29	30.2	182312	2	AP001655	AP001655 Homo sapi
23	c	29	30.2	185789	9	AC080092	AC080092 Homo sapi
24	c	29	30.2	190066	9	AC010127	AC010127 Homo sapi
25	c	29	30.2	191194	2	AP001654	AP001654 Homo sapi
26		29	30.2	204696	2	AC021673	AC021673 Homo sapi
27	c	28.8	30.0	153	9	HS60D6R	Z61939 H.sapiens C
28	c	28.8	30.0	138350	2	AC010801	AC010801 Homo sapi
29	c	28.8	30.0	142126	9	AL596266	AL596266 Homo sapi
30	c	28.8	30.0	170676	9	AC026258	AC026258 Homo sapi
31	c	28.8	30.0	183125	9	AP002848	AP002848 Homo sapi
32		28.8	30.0	189203	2	AC084114	AC084114 Homo sapi
33		28.8	30.0	195180	2	AL590674	AL590674 Homo sapi
34	c	28.8	30.0	195477	2	AC079196	AC079196 Homo sapi
35		28.8	30.0	221125	2	AP003329	AP003329 Oryza sat
36	c	28.6	29.8	953	9	AC082499	AC082499 Homo sapi
37	c	28.6	29.8	58522	9	AL445263	AL445263 Human DNA
38	c	28.6	29.8	125291	9	AP006046	AP006046 Homo sapi
39	c	28.6	29.8	161920	9	AP002502	AP002502 Homo sapi
40	c	28.6	29.8	168376	9	AP000826	AP000826 Homo sapi
41	c	28.6	29.8	168922	9	AP002007	AP002007 Homo sapi
42	c	28.6	29.8	172438	2	AL603710	AL603710 Mus muscu
43	c	28.6	29.8	175419	2	AP000788	AP000788 Homo sapi
44	c	28.6	29.8	181183	2	AL391866	AL391866 Homo sapi
45	c	28.4	29.6	187516	2	AC009648	AC009648 Homo sapi
	c		29.6	1117	5	AF109017	AF109017 Cactospiz

ALIGNMENTS

RESULT 1
HOMOVDR1
LOCUS   HOMOVDR1   96 bp   DNA   PRI   10-SEP-1998
DEFINITION Homo sapiens vitamin D receptor gene, exon 1d.
ACCESSION   AF080454
VERSION   AF080454.1   GI:3561037
KEYWORDS   .
SEGMENT   1 of 3
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 96)
  AUTHORS   Crofts,L.A., Hancock,M.S., Morrison,N.A. and Eisman,J.A.
  TITLE   Multiple promoters direct the tissue-specific expression of novel
            N-terminal variant human vitamin D receptor gene transcripts
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (18), 10529-10534 (1998)
  MEDLINE   98393678
REFERENCE   2 (bases 1 to 96)
  AUTHORS   Crofts,L.A., Hancock,M.S., Morrison,N.A. and Eisman,J.A.

```
TITLE       Direct Submission
JOURNAL     Submitted (27-JUL-1998) Bone and Mineral Research Program, Garvan
            Institute of Medical Research, 384 Victoria Street, Sydney, NSW
            2010, Australia
FEATURES             Location/Qualifiers
     source          1..96
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /map="12q13-14"
     exon            1..96
                     /gene="vitamin D receptor"
                     /note="1d"
                     /number=1
BASE COUNT       19 a     16 c     37 g     24 t
ORIGIN

Query Match             100.0%;  Score 96;  DB 9;  Length 96;
  Best Local Similarity 100.0%;
  Matches   96;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 gttccttcctctgtcggggcgccttggcatggaggaataagaaaggaggacgatt 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTTTCCTTCCTCTGTCGGGGCGGCGCCCCTTGGCATGGAGGAATAAGAAAGGAGGACGATT 60

Qy         61 ggctgtcgatggtgctcagaactgctggagtggagg 96
              ||||||||||||||||||||||||||||||||||||
Db         61 GGCTGTCGATGGTGCTCAGAACTGCTGGAGTGGAGG 96


RESULT  2
LOCUS       AF109063                1119 bp    DNA             VRT       09-MAY-1999
DEFINITION  Camarhynchus psittacula strain DRF11110 mitochondrial control
            region, partial sequence.
ACCESSION   AF109063
VERSION     AF109063.1  GI:4761187
KEYWORDS    .
SOURCE      Camarhynchus psittacula.
  ORGANISM  Camarhynchus psittacula
            Mitochondrion Camarhynchus psittacula
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Passeriformes; Fringillidae;
            Emberizinae; Camarhynchus.
REFERENCE   1  (bases 1 to 1119)
  AUTHORS   Sato,A., O'hUigin,C., Figueroa,F., Grant,P.R., Grant,B.R., Tichy,H.
            and Klein,J.
  TITLE     Phylogeny of Darwin's finches as revealed by mtDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (9), 5101-5106 (1999)
  MEDLINE   99238489
REFERENCE   2  (bases 1 to 1119)
  AUTHORS   Sato,A., O'hUigin,C., Figueroa,F., Tichy,H., Grant,P.R., Grant,B.R.
            and Klein,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-NOV-1998) Immungenetik, Max-Planck-Institut fur
            Biologie, Corrensstrasse 42, Tuebingen 72076, Germany
FEATURES             Location/Qualifiers
     source          1..1119
                     /organism="Camarhynchus psittacula"
                     /organelle="mitochondrion"
                     /strain="DRF11110"
                     /db_xref="taxon:87178"
                     /note="Darwin's finches"
     misc_feature    <1..>1119
                     /note="mitochondrial control region"
BASE COUNT      327 a    340 c    146 g    306 t
ORIGIN

Query Match              31.2%;  Score 30;  DB 5;  Length 1119;
  Best Local Similarity  64.3%;  Pred. No. 9.5;
  Matches   45;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy         10 ttctgtcggggcgccttggcatggaggaataagaaaggaggacgattggctgtcga 69
              |||  ||| |||| | |||  |||||  || |  || |||| | | ||  |||| ||
Db       1037 TTTTGTAAGGCGCTGTTTTGTGTGGGGTGGGGGAATAAGGGGGAGGAAAATATGGTGATGA 978

Qy         70 tggtgctcag 79
              ||  ||| ||
Db        977 TGATGGTTAG 968


RESULT  3
LOCUS       AF110424/c              1119 bp    DNA             VRT       11-MAY-1999
DEFINITION  Camarhynchus psittacula mitochondrial control region, partial
            sequence.
ACCESSION   AF110424
VERSION     AF110424.1  GI:4761207
KEYWORDS    .
SOURCE      Camarhynchus psittacula.
  ORGANISM  Camarhynchus psittacula
            Mitochondrion Camarhynchus psittacula
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Passeriformes; Fringillidae;
            Emberizinae; Camarhynchus.
REFERENCE   1  (bases 1 to 1119)
  AUTHORS   Sato,A., O'hUigin,C., Figueroa,F., Grant,P.R., Grant,B.R., Tichy,H.
            and Klein,J.
  TITLE     Phylogeny of Darwin's finches as revealed by mtDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (9), 5101-5106 (1999)
  MEDLINE   99238489
REFERENCE   2  (bases 1 to 1119)
  AUTHORS   Sato,A., O'hUigin,C., Figueroa,F., Tichy,H., Grant,P.R., Grant,B.R.
            and Klein,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-DEC-1998) Immungenetik, Max-Planck-Institut fur
            Biologie, Corrensstrasse 42, Tuebingen 72076, Germany
FEATURES             Location/Qualifiers
     source          1..1119
                     /organism="Camarhynchus psittacula"
                     /organelle="mitochondrion"
                     /strain="DRF113"
                     /db_xref="taxon:87178"
                     /note="Darwin's finches"
     misc_feature    <1..>1119
                     /note="mitochondrial control region"
BASE COUNT      327 a    340 c    146 g    306 t
ORIGIN

Query Match              31.2%;  Score 30;  DB 5;  Length 1119;
  Best Local Similarity  64.3%;  Pred. No. 9.5;
  Matches   45;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy         10 ttctgtcggggcgccttggcatggaggaataagaaaggaggacgattggctgtcga 69
              |||  ||| |||| | |||  |||||  || |  || |||| | | ||  |||| ||
Db       1037 TTTTGTAAGGCGCTGTTTTGTGTGGGGTGGGGGAATAAGGGGGAGGAAAATATGGTGATGA 978

Qy         70 tggtgctcag 79
              ||  ||| ||
Db        977 TGATGGTTAG 968


RESULT  4
LOCUS       HS228A9                95375 bp    DNA             PRI       12-DEC-1999
DEFINITION  Human DNA sequence from clone CTA-228A9 on chromosome 22q12.3-13.32
            Contains the PLA2G6 gene for cytosolic calcium-independent
            phospholipase A2, the 5' part of the gene for a novel protein
            similar to FAS-ligand associated factor 3 and Insulin receptor
            tyrosine kinase 53 kD substrate, ESTs, STSs, GSSs and a putative
            CpG island, complete sequence.
ACCESSION   HS228A9
VERSION     AL022322.1  GI:3236085
KEYWORDS    HTG; CpG island; FAS; Insulin receptor; phospholipase; PLA2G6;
            tyrosine kinase.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 95375)
  AUTHORS   Grafham,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 17, 1998 this sequence version replaced gi:3021233.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 22, constructed by the Sanger Centre Chromosome 22
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr22
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep CTA-228A9 is
            from the human BAC library described in U-J. Kim et al. (1996)
            Genomics 34, 213-218.
            VECTOR: pBACl08L
            This sequence is the entire insert of clone CTA-228A9 The true left
            end of clone CTA-447C4 is at 57390 in this sequence. The start of
            this sequence overlaps with sequence AL022311 The end of this
            sequence overlaps with sequence AL021977.
FEATURES             Location/Qualifiers
     source          1..95375
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="q12.3-13.32"
                     /clone="CTA-228A9"
                     /clone_lib="CTT9788K-A1"
     misc_feature    45..425
                     /note="match: STS: Em:B18032
                     match: GSS: Em:B18032"
     repeat_region   171..301
                     /note="FlAM_A repeat: matches 3. .133 of consensus"
     repeat_region   1372..1681
                     /note="AluSx repeat: matches 3. .312 of consensus"
     repeat_region   1800..1871
                     /note="8 copies 9 mer ggggtcat 87 conserved"
     repeat_region   3205..3259
                     /note="MIR repeat: matches 116. .173 of consensus"
     misc_feature    3265..4370
                     /note="CpG island"
                     /evidence=not_experimental
     repeat_region   3883..3938
                     /note="28 copies 2 mer gg 78 conserved"
     unsure          3891..4080
                     /note="Forced join"
     repeat_region   4756..5067
                     /note="AluSx repeat: matches 1. .311 of consensus"
     repeat_region   5237..5414
                     /note="MIR repeat: matches 71. .248 of consensus"
     repeat_region   5712..6011
                     /note="AluSg repeat: matches 1. .300 of consensus"
     repeat_region   6021..6194
                     /note="LiMA2 repeat: matches 6133. .6308 of consensus"
     repeat_region   6232..6715
                     /note="L1MC/D repeat: matches 5031. .5565 of consensus"
     repeat_region   6935..7228
                     /note="AluSx repeat: matches 5. .292 of consensus"
     repeat_region   7357..7537
                     /note="MSTB repeat: matches 1. .168 of consensus"
     repeat_region   7538..7845
                     /note="AluSq repeat: matches 1. .303 of consensus"
     repeat_region   7846..8061
                     /note="MSTB repeat: matches 168. .426 of consensus"
     repeat_region   8208..8845
                     /note="AluSx repeat: matches 1. .294 of consensus"
     repeat_region   8846..9141
                     /note="L1MB8 repeat: matches 5336. .6034 of consensus"
     repeat_region   9142..9291
                     /note="AluSq repeat: matches 1. .299 of consensus"
     repeat_region   9526..9818
                     /note="L1MB8 repeat: matches 6034. .6173 of consensus"
     repeat_region   10028..10320
                     /note="AluSx repeat: matches 1. .293 of consensus"
     repeat_region   10647..11028
                     /note="L1M4 repeat: matches 2135. .2490 of consensus"
     repeat_region   11029..11332
                     /note="AluSq repeat: matches 1. .304 of consensus"
     repeat_region   11333..11546
                     /note="L1M4 repeat: matches 2490. .3402 of consensus"
     gene            complement(join(11776. .11834,12719. .12835,13067. .13138,
                     complement(join(11776. .11834,12719. .12835,13067. .13138,
                     22508. .22569,22911. .22997,23789. .23864,25131. .25181))
     CDS             complement(join(<11776. .11834,12719. .12835,13067. .13138,
                     22508. .22569,22911. .22997,23789. .23864,25131. .25181))
                     /gene="dJ1039K5.3"
                     /note="this gene fragment and fragment -.2b are part of
                     the same gene
                     supported by FGENES and GENSCAN
                     match: proteins: Tr:043858 Tr:060437"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="bk228A9.2 (novel protein similar to FAS-ligand
                     associated factor 3 and Insulin receptor tyrosine kinase
                     53 kD substrate)"
                     /protein_id="CAB62940.1"
                     /db_xref="GI:6572199"
                     /translation="MAPEMDQFYRSTMAIYKSIMEQFNPALENLVYLGNNYLRAFHAL
                     SEAAEVYFSAIQKIGERALQSPTSQILGEILVQMSDTQRHLNSDLEVVQTFHGGLLQ
                     HMEKNTKLDMQFIKDSRQHYELEYRHRAANLEKCMSELWRMERKRDKNVRMKESVNR
                     LHAQMQAFVSESQRA"
     repeat_region   13259..13374
                     /note="MIR repeat: matches 128. .251 of consensus"
     misc_feature    13402..13839
                     /note="match: STS: Em:AA149345"
     repeat_region   13810..13890
                     /note="MIR repeat: matches 58. .138 of consensus"
     repeat_region   13910..14216
                     /note="AluSx repeat: matches 3. .309 of consensus"
     repeat_region   14926..15212
                     /note="AluJb repeat: matches 1. .312 of consensus"
     repeat_region   15225..15461
                     /note="MSTB repeat: matches 184. .426 of consensus"
     repeat_region   15462..15638
                     /note="LiME2 repeat: matches 5837. .6016 of consensus"
     repeat_region   15639..15699
                     /note="MER3 repeat: matches 12. .72 of consensus"
     repeat_region   15702..15907
                     /note="LiME2 repeat: matches 6077. .6164 of consensus"
     repeat_region   15907..16227
                     /note="AluJo repeat: matches 1. .307 of consensus"
     repeat_region   16230..16537
                     /note="AluSx repeat: matches 3. .310 of consensus"
     repeat_region   17291..17413
                     /note="AluJo repeat: matches 25. .153 of consensus"
     repeat_region   17414..17722
                     /note="AluYa8 repeat: matches 1. .301 of consensus"
```

```
repeat_region   17723..17898
                /note="AluJo repeat: matches 153..309 of consensus"
repeat_region   18009..18236
                /note="L1MC/D repeat: matches 5280..5515 of consensus"
repeat_region   18237..18537
                /note="L1MC/D repeat: matches 5515..5783 of consensus"
repeat_region   18538..18859
                /note="AluSx repeat: matches 1..301 of consensus"
repeat_region   18897..19148
                /note="L1MC/D repeat: matches 5515..5783 of consensus"
repeat_region   19149..19451
                /note="L1MC5 repeat: matches 7359..7520 of consensus"
repeat_region   19742..19985
                /note="AluYa8 repeat: matches 1..301 of consensus"
repeat_region   19986..20296
                /note="L2 repeat: matches 2140..2346 of consensus"
repeat_region   20297..20537
                /note="AluSx repeat: matches 3..312 of consensus"
repeat_region   20894..21204
                /note="L2 repeat: matches 2346..2744 of consensus"
repeat_region   21667..21815
                /note="AluSx repeat: matches 2..311 of consensus"
repeat_region   22240..22364
                /note="MIR repeat: matches 49..212 of consensus"
repeat_region   22645..23771
                /note="MIR repeat: matches 76..205 of consensus"
repeat_region   24399..24625
                /note="L1M3e repeat: matches -487..673 of consensus"
repeat_region   24675..24829
                /note="MIR repeat: matches 15..255 of consensus"
mRNA            complement(join(26151..26961,27160..27233,28143..28310,
                30183..30371,30867..35415,35565,37751..37914,
                41027..41105,42925..43086,44110..44218,47487..47669,
                49644..49740,54638..54825,57761..57944,60094..60309,
                83874..84143))
                /gene="PLA2G6"
                /note="match: cDNAs: Em:U51898 Em:AF064594 Em:U88624
                match: ESTs: Em:H10676 Em:AA432599 Em:AI568578 Em:W65048
                Em:H32591 Em:AA529527 Em:AA185698 Em:AA592727 Em:AI569276
                Em:AA895273 Em:AI663704 Em:R18691 Em:AA312471 Em:AA812478
                Em:AI261761 Em:H51147 Em:AI358312 Em:AA601449 Em:AI194003
                Em:AI040936 Em:AI138871 Em:H55254 Em:AI218245 Em:AA629376
                Em:AA745301 Em:AA719052 Em:AI015050 Em:AI635468
```

```
Query Match        31.2%;   Score 30;  DB 9;  Length 95375;
Best Local Similarity  64.3%;   Pred. No. 14;
Matches 45;  Conservative 0;  Mismatches 25;  Indels 0;  Gaps 0;

Qy      27  ggcatggagtggaggaataagaaaagaagcgagcgagtcgatggtcgttgctcagaactgct  86
            ||| || |||||  ||||||  |||  |||     | |   |   |   ||     || || |
Db   77084  GGGATGGAGGCTGGGAAAACGCGCCAGGAGCAATTCCCTGTCAATGGGGAAAAGCACTGTG  77143

Qy      87  ggagtgggg  96
            |||||||||
Db   77144  TGGGTGGAGG  77153
```

```
RESULT   5

LOCUS       AC006349   111937 bp    DNA       PRI    07-OCT-2000
DEFINITION  Homo sapiens PAC clone RP5-892G5 from 14q24.3, complete sequence.
ACCESSION   AC006349
VERSION     AC006349.3  GI:7243873
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 111937)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
```

```
REFERENCE   2 (bases 1 to 111937)
  AUTHORS   Hou,S., Courtney,L., Maupin,R., Nguyen,C. and Shafer,S.
  TITLE     The sequence of Homo sapiens PAC clone RP5-892G5
  JOURNAL   Unpublished
REFERENCE   3 (bases 1 to 111937)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JAN-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4 (bases 1 to 111937)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5 (bases 1 to 111937)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On Mar 15, 2000 this sequence version replaced gi:5001544.
COMMENT     -------------- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: saplens@watson.wustl.edu
            -------------- Summary Statistics
            Center project name: H_DJ0892G05
            --------------

            NOTICE: This sequence may not represent the entire insert of this
            clone. It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            This clone from chromosome 14 was provided by Dr. Pieter de Jong,
            Roswell Park Cancer Institute, Human Genetics Department, Elm and
            Carlton Streets, Buffalo NY 14263-0001 USA.

            SOURCE INFORMATION:
            This clone was derived from human PAC library RPCI-5, prepared by
            Pieter de Jong and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu) using the method described by
            Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from
            one male donor.
            The clone may be obtained either from Genome Systems, Inc.
            (http://www.genomesystems.com) or from Research Genetics, Inc.
            (http://www.resgen.com); or from Pieter de Jong.
            VECTOR: pCYPAC2
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP5-919J22, 200 bp overlap; the
            clone sequenced to the right is RP5-84708, 200 bp overlap. Actual
            start of this clone is at base position 124062 of RP5-919J22;
            actual end is at base position 18034 of RP5-84708.
FEATURES             Location/Qualifiers
     source          1..111937
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14q24.3"
                     /clone="RP5-892G5"
                     /clone_lib="RPCI-5"
     repeat_region   1..59
```

```
repeat_region   139..445        /rpt_family="Alu"
repeat_region   1269..1330      /rpt_family="Alu"
repeat_region   1348..1658      /rpt_family="MER1_type"
repeat_region   1663..1794      /rpt_family="Alu"
repeat_region   1806..1885      /rpt_family="Alu"
repeat_region   1886..2113      /rpt_family="MER1_type"
repeat_region   2578..2848      /rpt_family="Alu"
repeat_region   2857..2876      /rpt_family="Alu"
repeat_region   3159..3258      /rpt_family="(TTTA)n"
repeat_region   3259..3566      /rpt_family="MIR"
repeat_region   3567..3624      /rpt_family="Alu"
repeat_region   3971..4054      /rpt_family="MIR"
misc_feature    4100..4126      /note="Similar to EST H89528 (NID:g1080006) yw28b06.r1"
repeat_region   4111..4422      /rpt_family="(TTTA)n"
repeat_region   4467..4492      /rpt_family="Alu"
misc_feature    4476..4493      /note="(T)n"
misc_feature    4476..4495      /note="similar to EST AW087332 (NID:g6043137) xb48b01.x1"
repeat_region   4493..4576      /rpt_family="Alu"
misc_feature    4576            /note="similar to EST AW103750 (NID:g6074485) xd45f05.x1"
repeat_region   4785..4872      /rpt_family="Alu"
repeat_region   4785..4872      /rpt_family="MIR"
misc_feature    5009..5031      /note="similar to EST AI348438 (NID:g4085644) qo24f04.x1"
misc_feature    5012..5031      /note="similar to EST AI041813 (NID:g3281007) oy34a04.x1"
misc_feature    5012..5031      /note="similar to EST AA748084 (NID:g2788042) nx68g09.s1"
misc_feature    5013..5030      /note="similar to EST AI949616 (NID:g5741926) wp92g11.x1"
misc_feature    5014..5024      /note="match to EST AW450715 (NID:g6991491)"
misc_feature    5014..5024      /note="similar to EST AW452276 (NID:g6993052)"
misc_feature    5018..5031      /note="similar to EST AI862002 (NID:g9526109) tw36a10.x1"
misc_feature    5018..5031      /note="similar to EST AI799408 (NID:g5364880) tw56a11.x1"
misc_feature    5018..5031      /note="similar to EST AA723611 (NID:g2741318) ah24g08.s1"
repeat_region   5166..5311      /rpt_family="MIR"
repeat_region   5342..5442      /rpt_family="MIR"
repeat_region   5470..5523      /rpt_family="MIR"
repeat_region   7079..7101      /rpt_family="(GGGGA)n"
repeat_region   7798..7845      /rpt_family="(TC)n"
repeat_region   7845..7945      /rpt_family="MIR"
misc_feature    9716..9729      /note="similar to EST AW327264 (NID:g9797759) dq01c04.x1"

misc_feature    10186..10209    /note="similar to EST W73910 (NID:g13844207) zd56h08.r1"
misc_feature    10193..10209    /note="similar to EST AI799408 (NID:g5364880) tw56a11.x1"
misc_feature    10193..10209    /note="similar to EST AA207275 (NID:g1802768) zq55b07.r1"
misc_feature    10195..10210    /note="similar to EST R69222 (NID:g842739) yl39h06.r1"
misc_feature    10196..10210    /note="similar to EST H61278 (NID:g1014110) yu40e10.r1"
misc_feature    10387..10395    /note="match to EST AI640555 (NID:g4703664) wa15e02.x1"
repeat_region   10857..10886    /rpt_family="(CAAAA)n"
misc_feature    10967..11027    /note="match to EST L48868 (NID:g1254977)"
repeat_region   11184..11356    /rpt_family="MIR"
misc_feature    11338..13341    /note="CpG island (%GC=68.1, O/e=0.80, #CpGs=196)"
repeat_region   13905..13972    /rpt_family="L2"
repeat_region   14121..14236    /rpt_family="Alu"
repeat_region   14919..14980    /rpt_family="L2"
repeat_region   15040..15121    /rpt_family="MER1_type?"
repeat_region   15622..15906    /rpt_family="MIR"
repeat_region   16138..16319    /rpt_family="Alu"
repeat_region   16320..16616    /rpt_family="MIR"
repeat_region   16627..16929    /rpt_family="Alu"
repeat_region   17039..17133    /rpt_family="L2"
repeat_region   17134..17202    /rpt_family="MIR"
```

```
Query Match      31.0%;  Score 29.8;  DB 9;  Length 111937;
Best Local Similarity 58.4%;  Pred. No. 17;
Matches   52;  Conservative   0;  Mismatches   37;  Indels   0;  Gaps   0;

Qy      1  gtttccttcttctctgcgggggcgccctggccgtggagtcggaggaaataagaagaaaggaggcggatt   60
           ||| |||| || ||| | ||| || ||||| |||||||  |||||||||||||||| |||| |||
Db  75945  GTTACTTTGGTGTCTAGCGACGGGCAGGTTGGGAGAGAGTGAGGAGGAGGAATAAGGAATGACTCAAAT   76004

Qy     61  ggctgtcgatggtgctccagaaactgctgga   89
           ||| ||||||| || ||| ||| ||| ||| |||
Db  76005  GGATATGGAGTTTTCTTTTTGGGGTGATAGA   76033
```

```
RESULT   6
AC092439_0
WPCOMMENT
Sequence split into 4 fragments   LOCUS AC092439 Accession AC092439
```

Fragment Name	Begin	End
AC092439_0	1	110000
AC092439_1	100001	210000
AC092439_2	200001	310000
AC092439_3	300001	399218

```
LOCUS       AC092439   399218 bp   DNA   HTG   27-JUL-2001
DEFINITION  Homo sapiens chromosome UNK clone RP11-385C9, *** SEQUENCING IN
            PROGRESS ***, 22 unordered pieces.
ACCESSION   AC092439
VERSION     AC092439.1   GI:14595961
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 399218)
  AUTHORS    Waterston,R.H.
  TITLE      The sequence of Homo sapiens clone
  JOURNAL    Unpublished
REFERENCE   2  (bases 1 to 399218)
  AUTHORS    Waterston,R.H.
  TITLE      Direct Submission
  JOURNAL    Submitted (04-JUL-2001) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis,
             MO 63108, USA
COMMENT      -------------- Genome Center --------------
             Center: Washington University Genome Sequencing Center
             Center code: WUGSC
             Web site: http://genome.wustl.edu/gsc/index.shtml
             Center project name: H_NH0385C09
             -------------- Project Information --------------
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 22 contigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence
           * as soon as it is available and the accession number will
           * be preserved.
           *      1: 10000: contig of 10000 bp in length
           *  10000: 10100: gap of unknown length
           *  10100: 10100: contig of 10000 bp in length
           *  10100: 20100: gap of unknown length
           *  20100: 20201: contig of 10000 bp in length
           *  20201: 20201: gap of unknown length
           *  20201: 24628: contig of 4427 bp in length
           *  24628: 24727: gap of unknown length
           *  24727: 49260: contig of 24533 bp in length
           *  49260: 49361: gap of unknown length
           *  49361: 89060: contig of 39700 bp in length
           *  89060: 89161: gap of unknown length
           *  89161: 90956: contig of 1795 bp in length
           *  90956: 91055: gap of unknown length
           *  91055: 93082: contig of 2027 bp in length
           *  93082: 93183: gap of unknown length
           *  93183: 103575: contig of 10392 bp in length
           * 103575: 103674: gap of unknown length
           * 103674: 106604: contig of 2930 bp in length
           * 106604: 106705: gap of unknown length
           * 106705: 110034: contig of 3329 bp in length
           * 110034: 110133: gap of unknown length
           * 110133: 120034: contig of 10000 bp in length
           * 120034: 120133: gap of unknown length
           * 120133: 126029: contig of 5896 bp in length
           * 126029: 126130: gap of unknown length
           * 126130: 128973: contig of 2744 bp in length
           * 128973: 129074: gap of unknown length
           * 129074: 135624: contig of 6551 bp in length
           * 135624: 135725: gap of unknown length
           * 135725: 207836: contig of 72112 bp in length
           * 207836: 207937: gap of unknown length
           * 207937: 338618: contig of 130682 bp in length
           * 338618: 338719: gap of unknown length
           * 338719: 348718: contig of 10000 bp in length
           * 348718: 348819: gap of unknown length
           * 348819: 358818: contig of 10000 bp in length
           * 358818: 358919: gap of unknown length
           * 358919: 368918: contig of 10000 bp in length
           * 368918: 369019: gap of unknown length
           * 369019: 379018: contig of 10000 bp in length
           * 379018: 379119: gap of unknown length
           * 379119: 389118: contig of 10000 bp in length
           * 389118: 389219: gap of unknown length
           * 389219: 399218: contig of 10000 bp in length.
FEATURES             Location/Qualifiers
     source          1..399218
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
     misc_feature                 /chromosome="UNK"
                                  /clone="RP11-385C9"
     misc_feature    1..10000
                     /note="assembly_name:Contig1"
     misc_feature    10101..20100
                     /note="assembly_name:Contig10"
     misc_feature    20201..24627
                     /note="assembly_name:Contig11"
     misc_feature    24728..49260
                     /note="assembly_name:Contig13"
     misc_feature    49361..89060
                     /note="assembly_name:Contig14"
     misc_feature    89161..90955
                     /note="assembly_name:Contig15"
     misc_feature    91056..93082
                     /note="assembly_name:Contig16"
     misc_feature    93183..103574
                     /note="assembly_name:Contig17"
     misc_feature    103675..106604
                     /note="assembly_name:Contig18"
     misc_feature    106705..110033
                     /note="assembly_name:Contig19"
     misc_feature    110134..120133
                     /note="assembly_name:Contig2"
     misc_feature    120234..126129
                     /note="assembly_name:Contig20"
     misc_feature    126230..128973
                     /note="assembly_name:Contig21"
     misc_feature    129074..135624
                     /note="assembly_name:Contig22"
     misc_feature    135725..207836
                     /note="assembly_name:Contig23"
     misc_feature    207937..338618
                     /note="assembly_name:Contig24"
     misc_feature    338719..348718
                     /note="assembly_name:Contig3"
     misc_feature    348819..358818
                     /note="assembly_name:Contig4"
     misc_feature    358919..368918
                     /note="assembly_name:Contig5"
     misc_feature    369019..379018
                     /note="assembly_name:Contig6"
     misc_feature    379119..389118
                     /note="assembly_name:Contig7"
     misc_feature    389219..399218
                     /note="assembly_name:Contig9"
BASE COUNT   107809 a  92175 c  91936 g 105198 t   2100 others
ORIGIN
```

```
RESULT 7
AC026090/C
LOCUS       AC026090   172027 bp    DNA
DEFINITION  Homo sapiens chromosome 11 clone RP11-658K18, WORKING DRAFT
            SEQUENCE, 19 unordered pieces.
ACCESSION   AC026090
VERSION     AC026090.5  GI:9958310
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
Query Match             30.6%;  Score 29.4;  DB 2;  Length 110000;
Best Local Similarity   63.4%;  Pred. No. 24;
Matches   45;  Conservative   0;  Mismatches   26;  Indels   0;  Gaps   0;

Qy         17 gggcgccttggcatggagtggaggaataagaaaagacgattggctcgatggtgct 76
              |||||| ||| || ||| | |||||| |||| | | || |||| | || ||||||
Db      33011 GGGCGCCCAGGAGTATGGACTGAGGAAGACACAGAAGAGGGCCATGGGCTGGGGATGGATGGT 33070

Qy         77 cagaactgctg 87
              | |||||| |||
Db      33071 CGGCAGTCCTG 33081
```

```
REFERENCE   1  (bases 1 to 172027)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 172027)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Sep 1, 2000 this sequence version replaced gi:9838305.
```

```
            --------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0658K18
            ---------------- Summary Statistics ----------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 160096 bases at least Q40
            Consensus quality: 163968 bases at least Q30
            Consensus quality: 165626 bases at least Q20
            Insert size: 188000; agarose-fp
            Insert size: 171387; sum-of-contigs
            Quality coverage: 4.33 in Q20 bases; agarose-fp
            Quality coverage: 4.82 in Q20 bases; sum-of-contigs
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 19 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *      1      1526: contig of 1526 bp in length
            *   1527      1626: gap of unknown length
            *   1627      2975: contig of 1349 bp in length
            *   2976      3075: gap of unknown length
            *   3076      5073: contig of 1998 bp in length
            *   5074      5173: gap of unknown length
            *   5174      9317: contig of 4144 bp in length
            *   9318      9417: gap of unknown length
            *   9418     13195: contig of 3778 bp in length
            *  13196     13295: gap of unknown length
            *  13296     16482: contig of 3187 bp in length
            *  16483     16582: gap of unknown length
            *  16583     21965: contig of 5383 bp in length
            *  21966     22065: gap of unknown length
            *  22066     33167: contig of 11102 bp in length
            *  33168     33267: gap of unknown length
            *  33268     43162: contig of 9895 bp in length
            *  43163     43262: gap of unknown length
            *  43263     53157: contig of 9895 bp in length
            *  53158     53257: gap of unknown length
            *  53258     61965: contig of 8708 bp in length
            *  61966     62065: gap of unknown length
            *  62066     70847: contig of 8782 bp in length
            *  70848     70947: gap of unknown length
            *  70948     84297: contig of 13350 bp in length
            *  84298     84397: gap of unknown length
            *  84398    100019: contig of 15622 bp in length
            *  100020   100119: gap of unknown length
            *  100120   114785: contig of 14666 bp in length
            *  114786   114885: gap of unknown length
            *  114886   131872: contig of 16987 bp in length
```

```
FEATURES             Location/Qualifiers
     source          1..172027
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /clone="RP11-658K18"
     *   131873    131972: gap of unknown length
     *   131973    147618: contig of 15646 bp in length
     *   147619    147718: gap of unknown length
     *   147719    171498: contig of 23780 bp in length
     *   171499    171598: gap of unknown length
     *   171599    172027: contig of 429 bp in length.
     misc_feature    1..1526
                     /note="assembly_name:Contig14"
     misc_feature    1627..2975
                     /note="assembly_name:Contig17"
     misc_feature    3076..5073
                     /note="assembly_name:Contig18"
     misc_feature    5174..9317
                     /note="assembly_name:Contig19"
     misc_feature    9418..13195
                     /note="assembly_name:Contig20"
     misc_feature    13296..16482
                     /note="assembly_name:Contig21"
     misc_feature    16583..21965
                     /note="assembly_name:Contig22"
     misc_feature    22066..33167
                     /note="assembly_name:Contig23"
     misc_feature    33268..43162
                     /note="assembly_name:Contig24"
     misc_feature    43263..53157
                     /note="assembly_name:Contig25"
     misc_feature    53258..61965
                     /note="assembly_name:Contig26"
     misc_feature    62066..70847
                     /note="assembly_name:Contig27
                     vector_side:right"
                     clone_end:T7
     misc_feature    70948..84297
                     /note="assembly_name:Contig28"
     misc_feature    84398..100019
                     /note="assembly_name:Contig29"
     misc_feature    100120..114785
                     /note="assembly_name:Contig30"
     misc_feature    114886..131872
                     /note="assembly_name:Contig31"
     misc_feature    131973..147618
                     /note="assembly_name:Contig32"
     misc_feature    147719..171498
                     /note="assembly_name:Contig33"
     misc_feature    171599..172027
                     /note="assembly_name:Contig16"
BASE COUNT    53036 a  32800 c  32896 g  51491 t   1804 others
ORIGIN
```

```
RESULT      8
AC009412
LOCUS       AC009412       212827 bp    DNA             PRI       23-MAY-2001
DEFINITION  Homo sapiens clone RP11-425P5, complete sequence.
ACCESSION   AC009412
VERSION     AC009412.6  GI:14190769
KEYWORDS    HTG.
SOURCE      human.
```

```
Query Match        30.6%;  Score 29.4;  DB 2;  Length 172027;
Best Local Similarity   70.9%;  Pred. No. 25;
Matches   39;  Conservative   0;  Mismatches   16;  Indels   0;  Gaps   0;

Qy     29  catgggagtgggggaataaagaaaagggagcgattggcgtgtcgtggtcgtccagaact  83
           ||||||||||  | |||||||||||||||  |||||||| || |||||||| |||||| ||
Db  43074  CATGGTGGGCAGGATTAAGAAAAATTTGAGATTGGCTGGGCACCGTGCTCATGCCT  43020
```

```
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 212827)
AUTHORS     Waterston,R.H.
TITLE       The sequence of Homo sapiens clone
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 212827)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (21-AUG-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 212827)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (23-MAY-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On May 23, 2001 this sequence version replaced gi:13431187.
FEATURES             Location/Qualifiers
     source          1..212827
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-425P5"
BASE COUNT    58387 a   50075 c   49160 g   55205 t
ORIGIN
```

```
Query Match           30.6%;  Score 29.4;  DB 9;  Length 212827;
Best Local Similarity 63.4%;
Matches  45;  Conservative   0;  Mismatches  26;  Indels   0;  Gaps   0;

Qy         17 ggggcgccttggcatggagtggaggaataagaaaagagaagatggtgct 76
              ||| |||||||| |||||||||||||||||||| |||| |||||||| |
Db     196384 GGGGCCCAGAGAGTATGGACTGAGGAAGACACAAGAAGAAGACGGGGCC 196443
Qy         77 cagaactgctg 87
              |||| | |||
Db     196444 CGGCAGTCCTG 196454
```

```
RESULT  9
LOCUS       AC080964      854 bp    DNA    HTG    28-SEP-2000
DEFINITION  Giardia intestinalis clone AJ3202 strain WB-C6, LOW-PASS SEQUENCE
            SAMPLING.
ACCESSION   AC080964
VERSION     AC080964.1  GI:10355770
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Giardia intestinalis.
ORGANISM    Giardia intestinalis
            Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
REFERENCE   1  (bases 1 to 854)
AUTHORS     Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.O., Kim,U.,
            Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
TITLE       Giardia: a model for ancient eukaryotic genome analysis
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 854)
AUTHORS     McArthur,A.G., Morrison,H.G., Nixon,J., Eakin,N.O., Kim,U.,
            Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
TITLE       Direct Submission
JOURNAL     Submitted (28-SEP-2000) Josephine Bay Paul Center for Comparative
            Molecular Biology and Evolution, Marine Biological Laboratory, 7
            MBL Street, Woods Hole, MA 02543-1015, USA
COMMENT     NOTE: This record contains 1 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
```

```
FEATURES             Location/Qualifiers
     source          1..854: contig of 854 bp in length.
                     /organism="Giardia intestinalis"
                     /strain="WB-C6"
                     /db_xref="taxon:5741"
                     /clone="AJ3202"
BASE COUNT    253 a   164 c   231 g   204 t   2 others
ORIGIN
```

```
Query Match           30.4%;  Score 29.2;  DB 2;  Length 854;
Best Local Similarity 58.3%;
Matches  49;  Conservative   1;  Mismatches  34;  Indels   0;  Gaps   0;

Qy         12 ctgtcggggcgccttggccatggagtggaggaataagaaaaggggaggaagatggtgct 71
              || | |||||| ||| |||||| || |||||||||| ||| || |||||| ||| || |
Db        170 CTCTCGGGGGTGTCTCGAATAAGGGGTCGAGTATTCACCAGAGTATAGAGATGTGCTGTGAATG 229
Qy         72 gtgtctcagaactgctggagtggag 95
              ||| ||||| ||||| | || |||
Db        230 CAGCGCGASAATCGCAGAGTTGGAG 253
```

```
RESULT  10
LOCUS       AC004877     128361 bp    DNA    PRI    21-DEC-1999
DEFINITION  Homo sapiens PAC clone RP4-751H13 from 7q35-qter, complete
            sequence.
ACCESSION   AC004877
VERSION     AC004877.1  GI:3638954
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 128361)
AUTHORS     Leonard,S., Graves,T. and Strowmatt,C.
TITLE       The sequence of Homo sapiens PAC clone RP4-751H13
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 128361)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 128361)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (19-SEP-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4  (bases 1 to 128361)
AUTHORS     Waterston,R.
TITLE       Direct Submission
JOURNAL     Submitted (21-DEC-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Sep 19, 1998 this sequence version replaced gi:3213120.
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watson.wustl.edu
            Summary Statistics
            Center project name: H_DJ0751H13
            -----------------
            Genome Center
```

```
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
```

```
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
```

between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/GTB/CHR7, send
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP4-811N16, 200 bp overlap.
Actual start of this clone is at base position 1 of RP4-751H13;
actual end is at 128361 of RP4-751H13.

```
FEATURES             Location/Qualifiers
source               1..128361
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q35-qter"
                     /clone="RP4-751H13"
                     /clone_lib="unknown"
repeat_region        20..155
                     /rpt_family="Alu"
repeat_region        860..941
                     /rpt_family="Alu"
repeat_region        2531..2885
                     /rpt_family="MER1_type?"
repeat_region        2924..3246
                     /rpt_family="Retroviral"
repeat_region        3300..3727
                     /rpt_family="Alu"
repeat_region        4146..4287
                     /rpt_family="Retroviral"
repeat_region        5253..5728
                     /rpt_family="L2"
repeat_region        5933..6243
                     /rpt_family="MaLR"
repeat_region        7787..7870
                     /rpt_family="Alu"
repeat_region        7798..8145
                     /rpt_family="MIR"
misc_feature         8259..8311
                     /note="match to EST AA084720 (NID:g1626811) zn06f02.s1"
repeat_region        9100..9426
                     /rpt_family="L2"
misc_feature         9773..10086
                     /note="match to EST AA080867 (NID:g1625560) zn06f02.r1"
repeat_region        complement(9947..10370)
                     /rpt_family="Alu"
misc_feature         9947..10370
                     /note="match to EST AA828832 (NID:g2901931) od80c03.s1"
misc_feature         10066..10388
                     /note="match to EST AA828832 (NID:g2901931) od80c03.s1"
repeat_region        10297..10734
                     /rpt_family="MaLR"
repeat_region        11605..12220
                     /rpt_family="L1"
repeat_region        12222..12615
                     /rpt_family="Retroviral"
repeat_region        12617..12755
                     /rpt_family="L1"
repeat_region        12768..12890
                     /rpt_family="L1"
repeat_region        12891..13196
                     /rpt_family="MIR"
repeat_region        13197..13309
                     /rpt_family="Alu"
repeat_region        13314..13618
                     /rpt_family="MIR"
repeat_region        13693..14004
                     /rpt_family="Alu"
repeat_region        14040..14162
                     /rpt_family="Alu"
repeat_region        15625..15591
                     /rpt_family="MIR"
repeat_region        15931
                     /rpt_family="Alu"
misc_feature         16550..16702
                     /note="match to EST AA939299 (NID:g3099212) o178g08.s1"
repeat_region        18090..19036
                     /rpt_family="L1"
repeat_region        19479..19798
                     /rpt_family="L2"
misc_feature         19479..19656
                     /note="match to EST AA923119 (NID:g3070428) ok91e12.s1"
misc_feature         19817..20311
                     /note="similar to EST H89462 (NID:g1079892)"
misc_feature         19860..21428
                     /note="match to EST AA625192 (NID:g2537577) af67h06.r1"
misc_feature         complement(20009..21523)
                     /note="CpG island (%GC=70.4, o/e=0.70, #CpGs=171)"
gene                 complement(19861..20173)
                     /gene="WUGSC:H_DJ0751H13.3"
CDS                  complement(20009..21523)
                     /gene="WUGSC:H_DJ0751H13.3"
                     /note="zinc finger-like; similar to P52742 (PID:g1731411);
                     H_DJ0751H13.3"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAC36300.1"
                     /db_xref="GI:3638956"
                     /translation="MIRKVKVEDEDQEAEEEVEWPQHLSLLPSFPFPAPDLGHLAAAYK
                     LEPGAPGALSGLALSGWGPMPEKPYGCGGECERRFRDQLTLRLHQRLHRGEGPCACPDC
                     GRSFTQRAHMLLHQRSHRGERPFPCPSECDKRFSKKAHLTRHLRTHTGERPYCAQCARSF
                     RFSQKIHLGSHQKTHTGERPFPCTECEKRFRRKTHLIRHQRIHTGERPYCAQCARSF
                     THKQHLVRHQRVHQTAGPARPSPDSSASPHSTAPSPTPSFPGPKFPACSDCGLSFGWK
                     KNLATHQCLHRSEGRFPGCDECALGATVDAPAAKPLASAPGGPCGGPCSSDPVVPQRAP
                     SGERSFFCPDCGRGFSHGQHLARHHPRVHTGERFPACTQCCDRRFGSRPNLVAHSRAHSG
                     ARPFACAQCGRRFSRKSHLGRHQAVHTGSRPHACAVCARSFSSKTNLVRHQAIHTGSR
                     PFSCPQCGKSFSRKTHLVRHQLIHGEAAHAAPDAALAAPAWSAPPEVAPPLFF"
misc_feature         complement(20014..20272)
                     /note="WUGSC:H_DJ0751H13.3"
misc_feature         complement(20059..20239)
                     /note="match to EST AA771702 (NID:g2823485) ai132a10.s1"
misc_feature         complement(20097..20319)
                     /note="match to EST AA923119 (NID:g3070428) ok91e12.s1"
misc_feature         20176..20322
                     /note="match to EST AA885789 (NID:g3000897) oj35a06.s1"
misc_feature         20298..20740
                     /note="match to EST AA944411 (NID:g3030812) oe79b04.s1"
misc_feature         20298..20740
                     /note="similar to Mus musculus EST AA021873 (NID:g1485629)
                     mh85fi2.r1"
misc_feature         20534..20724
                     /note="match to EST AA044857 (NID:g1523196) zf54d12.r1"
misc_feature         20819..21219
                     /note="match to EST AA757828 (NID:g2805691) zg44d02.s1"
repeat_region        22351..22486
                     /rpt_family="MIR"
repeat_region        22848..23146
                     /rpt_family="MIR"
repeat_region        23679..23794
                     /rpt_family="Alu"
```

```
RESULT 11

Qy          12  ctgtcggggcgccttggcatggaggtgaggaataagaaaagggaggcgcgtggccgtcgatg 71
                |||||||||||| |||||| ||| ||||||||||||||||||||||||| ||| |||  |
Db 113856  CTGTGGGGTGCTCTCTGGGGGAGGAAGTGAAGGACAAAAGGAAAAAGGGAGATTGGTTTCCACTA 113797

Qy          72  gtgctc 77
                || ||
Db 113796  CTCATC 113791

                Query Match            30.4%;  Score 29.2;  DB 9;  Length 128361;
                Best Local Similarity  65.2%;  Pred. No. 28;
                Matches    43;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy          12  ctgtcggggcgccttggcatggaggtgaggaataagaaaagggaggcgcgtggccgtcgatg 71
```

```
repeat_region          23773..23846          /rpt_family="MIR"
repeat_region          24494..24982          /rpt_family="L2"
repeat_region          24983..25271          /rpt_family="L2"
repeat_region          25272..25616          /rpt_family="Alu"
repeat_region          26693..29260          /rpt_family="L2"
misc_feature                                 /note="CpG_island (%GC=73.8, o/e=0.90, #CpGs=195)"
```

```
LOCUS       AC040944  150771 bp  DNA  HTG  12-MAY-2000
DEFINITION  Homo sapiens chromosome 4 clone RP11-194H16 map 4, WORKING DRAFT
            SEQUENCE, 18 unordered pieces.
ACCESSION   AC040944
VERSION     AC040944.2  GI:7770555
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 150771)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 4, clone RP11-194H16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 150771)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,M., Bastien,V., Beda,F.,
            Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Lakocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 12, 2000 this sequence version replaced gi:7534143.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------------------------------------------------------------
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
```

```
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            --------------------------------------------------------------
            -- Project Information
            Center project name: L9006
            Center clone name: 194_H_16
            --------------------------------------------------------------
            -- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 139261 bases at least Q40
            Consensus quality: 144895 bases at least Q30
            Consensus quality: 147331 bases at least Q20
            Insert size: 157000; agarose-fp
            Insert size: 149071; sum-of-contigs
            Quality coverage: 3.8 in Q20 bases; agarose-fp
            Quality coverage: 4.0 in Q20 bases; sum-of-contigs
            --------------------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *          1  1289: contig of 1289 bp in length
            *       1290  1389: gap of 100 bp
            *       1390  2512: contig of 1123 bp in length
            *       2513  2612: gap of 100 bp
            *       2613  4453: contig of 1841 bp in length
            *       4454  4553: gap of 100 bp
            *       4554  9223: contig of 4670 bp in length
            *       9224  9323: gap of 100 bp
            *       9324 13286: contig of 3963 bp in length
            *      13287 13386: gap of 100 bp
            *      13387 17467: contig of 4081 bp in length
            *      17468 17567: gap of 100 bp
            *      17568 22638: contig of 5071 bp in length
            *      22639 22738: gap of 100 bp
            *      22739 27698: contig of 4960 bp in length
            *      27699 27798: gap of 100 bp
            *      27799 33999: contig of 6201 bp in length
            *      34000 34099: gap of 100 bp
            *      34100 39232: contig of 5133 bp in length
            *      39233 39332: gap of 100 bp
            *      39333 47859: contig of 8527 bp in length
            *      47860 47959: gap of 100 bp
            *      47960 55328: contig of 7369 bp in length
            *      55329 55428: gap of 100 bp
            *      55429 63346: contig of 7918 bp in length
            *      63347 63446: gap of 100 bp
            *      63447 75602: contig of 12156 bp in length
            *      75603 75702: gap of 100 bp
            *      75703 89792: contig of 14090 bp in length
            *      89793 89892: gap of 100 bp
            *      89893 106216: contig of 16324 bp in length
            *     106217 106316: gap of 100 bp
            *     106317 125689: contig of 19373 bp in length
            *     125690 125789: gap of 100 bp
            *     125790 150771: contig of 24982 bp in length
FEATURES             Location/Qualifiers
     source          1..150771
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-194H16"
                     /note="RP11-194H16 Human Male BAC"
                     /clone_lib="RPCI-11"
     misc_feature    1..1289
                     /note="assembly_fragment"
     misc_feature    1390..2512
                     /note="assembly_fragment"
     misc_feature    2613..4453
                     /note="assembly_fragment"
```

```
misc_feature    4554..9223
                /note="assembly_fragment"
                clone_end:SP6
                vector_side:left"
misc_feature    9324..13286
                /note="assembly_fragment"
misc_feature    13387..17467
                /note="assembly_fragment"
misc_feature    17568..22638
                /note="assembly_fragment"
misc_feature    22739..27698
                /note="assembly_fragment"
misc_feature    27799..33999
                /note="assembly_fragment"
misc_feature    34100..34932
                /note="assembly_fragment"
misc_feature    33933..47859
                /note="assembly_fragment"
misc_feature    47960..55328
                /note="assembly_fragment"
misc_feature    55429..63346
                /note="assembly_fragment"
misc_feature    63447..75602
                /note="assembly_fragment"
misc_feature    75703..89792
                /note="assembly_fragment"
misc_feature    89993..106216
                /note="assembly_fragment"
misc_feature    106317..125689
                /note="assembly_fragment"
misc_feature    125790..150771
                /note="assembly_fragment"
                clone_end:T7
                vector_side:left"

BASE COUNT     38662 a   36806 c   36485 g   37112 t   1706 others
ORIGIN
```

```
RESULT 12
AL596186

Query Match
Best Local Similarity  30.4%;
Matches 46;  Conservative 0;  Mismatches 28;  Indels 0;  Gaps 0;
Score 29.2;  DB 2;  Length 150771;
62.2%;  Pred. No. 28;

Qy        1 gttccttccttctgtcgggggcgccttggcatggagtggaggaataagagaaaggagcgatt 60
            ||| ||| |||||  | ||||||| | |||  ||  | |  |||| |   |||||||||||||
Db    98703 GTTCCCTGCCTCTCCTGCTTGTCCTCCACTCGCCTTTTGGTGAAGAGGAGAAAGAAGAGCCCTT 98762

Qy       61 ggctgtcgatggtg 74
            ||| | |||||
Db    98763 GGCAGGGCATGGTG 98776
```

```
LOCUS       AL596186   154998 bp   DNA   HTG   20-JUL-2001
DEFINITION  Homo sapiens chromosome 6 clone RP11-48B6, *** SEQUENCING IN
            PROGRESS ***, 17 unordered pieces.
ACCESSION   AL596186
VERSION     AL596186.1  GI:15020794
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Homo sapiens
            human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 154998)
AUTHORS     Sims,S.
TITLE       Direct Submission
JOURNAL     Submitted (20-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
            requests: clonerequest@sanger.ac.uk
COMMENT
            Center: Sanger Centre
            Center code: SC
            Center: Genome Center
```

```
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            Center project name: bA4B6
            --------------------  Project Information
            --------------------  Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 141876 bases at least Q40
            Consensus quality: 147231 bases at least Q30
            Consensus quality: 150692 bases at least Q20
            Insert size: 153398; sum-of-contigs
            Insert size: 183237; 20.3% error; agarose-fp
            Quality coverage: 3.44x in Q20 bases; sum-of-contigs Quality
            coverage: 2.90x in Q20 bases; agarose-fp
            *
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *        1   15837: contig of 15837 bp in length
            *    15838   15937: gap of     100 bp
            *    15938   25580: contig of  9643 bp in length
            *    25581   25680: gap of     100 bp
            *    25681   29702: contig of  4022 bp in length
            *    29703   29802: gap of     100 bp
            *    29803   39318: contig of  9516 bp in length
            *    39319   39418: gap of     100 bp
            *    39419   54030: contig of 14612 bp in length
            *    54031   54130: gap of     100 bp
            *    54131   57818: contig of  3688 bp in length
            *    57819   57918: gap of     100 bp
            *    57919   77013: contig of 19095 bp in length
            *    77014   77113: gap of     100 bp
            *    77114   80109: contig of  2996 bp in length
            *    80110   80209: gap of     100 bp
            *    80210   93780: contig of 13571 bp in length
            *    93781   93880: gap of     100 bp
            *    93881  100480: contig of  6600 bp in length
            *   100481  100580: gap of     100 bp
            *   100581  104869: contig of  4289 bp in length
            *   104870  104969: gap of     100 bp
            *   104970  132041: contig of 27072 bp in length
            *   132042  132141: gap of     100 bp
            *   132142  134949: contig of  2808 bp in length
            *   134950  135049: gap of     100 bp
            *   135050  138690: contig of  3641 bp in length
            *   138691  138790: gap of     100 bp
            *   138791  148210: contig of  9420 bp in length
            *   148211  148310: gap of     100 bp
            *   148311  152396: contig of  4086 bp in length
            *   152397  152496: gap of     100 bp
            *   152497  154998: contig of  2502 bp in length
FEATURES             Location/Qualifiers
     source          1..154998
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-48B6"
                     /clone_lib="RPCI-11.1"
     misc_feature    1..15837
                     /note="assembly_fragment:01520
                     fragment_chain:1"
     misc_feature    15938..25580
                     /note="assembly_fragment:00242
                     fragment_chain:1"
                     clone_end:T7
                     vector_side:left"
     misc_feature    25681..29702
                     /note="assembly_fragment:00242
                     fragment_chain:1"
```

```
RESULT 13

LOCUS       AL354977    163447 bp    DNA           PRI       27-NOV-2000
DEFINITION  Human DNA sequence from clone RP11-509J21 on chromosome 9, complete
            sequence.
ACCESSION   AL354977
VERSION     AL354977.10  GI:10803246
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163447)
  AUTHORS   Lloyd,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequests@sanger.ac.uk
COMMENT     On Oct 16, 2000 this sequence version replaced gi:10086106.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:; EMBL; Sw:; SWISSPROT; Tr:; TREMBL; Wp:; WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-509J21 is from the library RPCI-11.2 constructed at the
            Roswell Park Cancer Institute by the group of Pieter de Jong. For
            further details see http://bacpac.med.buffalo.edu/
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-509J21 The true
            left end of clone RP11-661K19 is at 114307 in this sequence. The
            true right end of clone RP11-326A8 is at 32194 in this sequence.
FEATURES             Location/Qualifiers
     source          1..163447
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-509J21"
                     /clone_lib="RPCI-11.2"
     repeat_region   1..547
                     /note="L1M4 repeat: matches 4282..4882 of consensus"
     repeat_region   1664..1739
                     /note="AluJ/FLAM repeat: matches 6..81 of consensus"
     repeat_region   1819..1872
                     /note="L2 repeat: matches 2345..2474 of consensus"
     repeat_region   2942..3214
                     /note="Alu repeat: matches 243..296 of consensus"
     repeat_region   3581..3702
                     /note="AluSx repeat: matches 39..312 of consensus"
     repeat_region   3999..4195
                     /note="AluJo/FRAM repeat: matches 185..295 of consensus"
     repeat_region   4258..4550
                     /note="MIR repeat: matches 45..262 of consensus"
     repeat_region   4797..5012
                     /note="AluY repeat: matches 1..293 of consensus"
     repeat_region   5026..5558
                     /note="AluSg/x repeat: matches 86..300 of consensus"
     repeat_region   5978..6102
                     /note="L2 repeat: matches 1710..2278 of consensus"
     repeat_region   6113..6349
                     /note="L2 repeat: matches 2345..2474 of consensus"
     repeat_region   6560..6861
                     /note="MIR repeat: matches 11..259 of consensus"
     repeat_region   8760..8975
                     /note="AluSg repeat: matches 1..302 of consensus"
     repeat_region   9395..9659
                     /note="AluJb repeat: matches 72..287 of consensus"
     repeat_region   10786..11196
                     /note="AluSx repeat: matches 1..284 of consensus"
     repeat_region   11402..11528
                     /note="MLT1J repeat: matches 1..414 of consensus"
     repeat_region   11632..11659
                     /note="MIR repeat: matches 1..149 of consensus"
                     /note="MIR repeat: matches 20..149 of consensus"
                     /note="14 copies 2 mer ac 100% conserved"

BASE COUNT     40488 a   37626 c   37687 g   37585 t   1612 others
ORIGIN
            vector_side:right

     misc_feature    29803..39318
                     /note="assembly_fragment:01017"
                     /note="fragment_chain:2"
     misc_feature    39419..54030
                     /note="assembly_fragment:01555"
                     /note="fragment_chain:2"
     misc_feature    54131..57818
                     /note="assembly_fragment:00630"
                     /note="fragment_chain:2"
     misc_feature    57919..77013
                     /note="assembly_fragment:01696"
                     /note="fragment_chain:3"
     misc_feature    77114..80109
                     /note="assembly_fragment:00740"
                     /note="fragment_chain:3"
     misc_feature    80210..93780
                     /note="assembly_fragment:00028"
                     /note="fragment_chain:3"
     misc_feature    93881..100480
                     /note="assembly_fragment:00082"
                     /note="fragment_chain:4"
     misc_feature    100581..104869
                     /note="assembly_fragment:01267"
                     /note="fragment_chain:4"
     misc_feature    104970..132041
                     /note="assembly_fragment:00914"
                     /note="fragment_chain:5"
     misc_feature    132142..134949
                     /note="assembly_fragment:00553"
                     /note="fragment_chain:5"
     misc_feature    135050..138690
                     /note="assembly_fragment:00469"
     misc_feature    138791..148210
                     /note="assembly_fragment:00726"
     misc_feature    148311..152396
                     /note="assembly_fragment:01015"
     misc_feature    152497..154998
                     /note="assembly_fragment:01193"
     misc_feature    ..
                     /note="assembly_fragment:00561"
                     /note="clone_end:SP6"
```

```
Query Match 30.4%;  Score 29.2;  DB 2;  Length 154998;
Best Local Similarity 62.2%;  Pred. No. 28;
Matches 46;  Conservative 0;  Mismatches 28;  Indels 0;  Gaps 0;

Qy      1 gttccttcctctgtcgggcgcgccttggcatggagtggaggaataagaaagagggcgatt 60
          ||||| || || |||     |||||||||||| |||||||||||||||||||||||
Db  33879 GTTCCCTGCCGCCTCTCCTTGTCCTCACTGGCTTTTGGTGAAGAGGAAAAAGAAGCCCTT 33938

Qy     61 ggctgtcgatggtg 74
          ||||| |||||||
Db  33939 GGCAGGGCATGGTG 33952
```

```
repeat_region   11729..11921
                /note="MIR repeat: matches 31..221 of consensus"
repeat_region   11943..12062
                /note="L1ME3A repeat: matches 5991..6113 of consensus"
repeat_region   12063..12621
                /note="L1ME3A repeat: matches 1..553 of consensus"
repeat_region   12622..12672
                /note="MLT2D repeat: matches 1..553 of consensus"
repeat_region   12742..13045
                /note="L1ME3A repeat: matches 6113..6164 of consensus"
repeat_region   13420..13511
                /note="AluSg repeat: matches 1..305 of consensus"
repeat_region   13692..20142
                /note="L1MC2 repeat: matches 6218..6314 of consensus"
repeat_region   20151..20339
                /note="L1PA7 repeat: matches 8..6139 of consensus"
repeat_region   20353..20841
                /note="L1MC2 repeat: matches 6044..6234 of consensus"
repeat_region   20842..21140
                /note="Tigger3b repeat: matches 701..1205 of consensus"
repeat_region   21141..21896
                /note="AluJb repeat: matches 1..288 of consensus"
repeat_region   21915..22265
                /note="L1MC2 repeat: matches 5643..6004 of consensus"
repeat_region   22435..22736
                /note="AluSg repeat: matches 1..303 of consensus"
repeat_region   23083..23379
                /note="AluSx repeat: matches 3..297 of consensus"
repeat_region   23984..24170
                /note="MIR repeat: matches 49..244 of consensus"
repeat_region   25443..25466
                /note="MER58C repeat: matches 4..89 of consensus"
repeat_region   25992..26189
                /note="AluSg/x repeat: matches 170..293 of consensus"
repeat_region   27058..27825
                /note="MIR repeat: matches 3..262 of consensus"
repeat_region   27781..27928
                /note="L2 repeat: matches 2099..2545 of consensus"
repeat_region   28125..28584
                /note="MIR repeat: matches 49..96 of consensus"
repeat_region   28661..28889
                /note="L2 repeat: matches 1992..2489 of consensus"
repeat_region   29272..29429
                /note="L2 repeat: matches 1643..1880 of consensus"
repeat_region   29485..29731
                /note="MER5B repeat: matches 1..175 of consensus"
repeat_region   31630..32055
                /note="MIR repeat: matches 3..262 of consensus"
repeat_region   32445..32490
                /note="L2 repeat: matches 2545..2545 of consensus"
repeat_region   33177..33580
                /note="MER20 repeat: matches 2..218 of consensus"
repeat_region   34501..34800
                /note="MLT1B repeat: matches 1..390 of consensus"
repeat_region   34910..35029
                /note="AluY repeat: matches 1..299 of consensus"
repeat_region   35051..36193
                /note="L2 repeat: matches 1931..2053 of consensus"
repeat_region   36596..36913
                /note="MER69B repeat: matches 1..1225 of consensus"
repeat_region   37179..37473
                /note="L1MD repeat: matches 977..1287 of consensus"
repeat_region   37497..37538
                /note="AluJb repeat: matches 1..307 of consensus"
repeat_region   37563..37976
                /note="21 copies 2 mer aa 83% conserved"
repeat_region   3797..38328
                /note="MSTA repeat: matches 1..426 of consensus"
repeat_region   38341..38696
                /note="MSTA repeat: matches 1..426 of consensus"
repeat_region   38762..38914
                /note="L1MEc repeat: matches 1639..1982 of consensus"

repeat_region   39048..40267
                /note="AluJo/FRAM repeat: matches 150..302 of consensus"
repeat_region   40296..41553
                /note="L1MEc repeat: matches 2189..3119 of consensus"
repeat_region   41554..41851
                /note="L2 repeat: matches 433..1770 of consensus"
repeat_region   41852..42145
                /note="AluJb repeat: matches 1..312 of consensus"
repeat_region   43119..43335
                /note="L2 repeat: matches 1770..2050 of consensus"
repeat_region   43646..44014
                /note="MIR repeat: matches 6..262 of consensus"
repeat_region   44015..44395
                /note="MSTA repeat: matches 1..375 of consensus"
repeat_region   44396..44448
                /note="THEIC repeat: matches 1..371 of consensus"
repeat_region   44572..45137
                /note="MSTA repeat: matches 375..426 of consensus"
repeat_region   46232..46445
                /note="L1PA4 repeat: matches 5575..6146 of consensus"
repeat_region   46771..47130
                /note="MIR repeat: matches 3..229 of consensus"
repeat_region   47651..47797
                /note="THEIB repeat: matches 1..364 of consensus"
repeat_region   48421..48768
                /note="AluJb repeat: matches 1..299 of consensus"
repeat_region   48769..49067
                /note="L2 repeat: matches 1867..2216 of consensus"
repeat_region   49068..49438
                /note="AluY repeat: matches 1..299 of consensus"
repeat_region   49670..49838
                /note="L2 repeat: matches 1500..1867 of consensus"
repeat_region   50426..50744
                /note="L2 repeat: matches 1232..1436 of consensus"
repeat_region   50973..51307
                /note="AluY repeat: matches 1..311 of consensus"
repeat_region   ........
                /note="HERVL repeat: matches 3569..3907 of consensus"
```

```
RESULT 14
LOCUS       AL353577   174476 bp    DNA       HTG    22-JUL-2001
DEFINITION  Homo sapiens chromosome 9 clone RP11-661K19, *** SEQUENCING IN
            PROGRESS ***, 3 unordered pieces.
ACCESSION   AL353577
VERSION     AL353577.19  GI:15021133
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174476)
AUTHORS     Mashreghi-Mohammadi,M.
TITLE       Direct Submission
JOURNAL     Submitted (21-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
```

```
Query Match     30.4%;  Score 29.2;  DB 9;  Length 163447;
Best Local Similarity   57.8%;   Pred. No. 29;   Indels 0;   Gaps 0;
Matches  52;  Conservative  0;  Mismatches 38;

Qy      2 ttccttcttctgtcggggcgccttggcatggagtggaggaataagaaaaggggcgattg 61
          ||||  ||||  ||  || | || | |  | | ||||  | || | |||||| | |||
Db 149995 TGTGCATCTATACTCCTGGCTTCTCGGACACGGTCGAGGTGGGAGAATCACATGAGCCAAAG 150054

Qy     62 gctgtcgatggtgctcagaactgctggagt 91
          | ||| || ||||| ||  |  ||||| | |
Db 150055 GAATTCGAGGCTCCATTGAACCACTGCCACT 150084
```

COMMENT   On Jul 25, 2001 this sequence version replaced gi:14586073.
           -------------------- Genome Center --------------------
           Center: Sanger Centre
           Center code: SC
           Web site: http://www.sanger.ac.uk
           Contact: humquery@sanger.ac.uk
           ----------------- Project Information -----------------
           Center project name: bA661K19
           ------------------ Summary Statistics -----------------
           Assembly program: XGAP4; version 4.5
           Sequencing vector: M13; M7815; 6% of reads
           Sequencing vector: plasmid; L08752; 93% of reads
           Chemistry: Dye-terminator Big Dye; 100% of reads
           Consensus quality: 174016 bases at least Q40
           Consensus quality: 174193 bases at least Q30
           Consensus quality: 174245 bases at least Q20
           Insert size: 174276; sum-of-contigs
           Insert size: 168504; 5.9% error; agarose-fp
           Quality coverage: 9.21x in Q20 bases; agarose-fp
           coverage: 10.11x in Q20 bases; agarose-fp

           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 3 contigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence
           * as soon as it is available and the accession number will
           * be preserved.

FEATURES   Location/Qualifiers
     source   1..174476
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-661K19"
                     /clone_lib="RPCI-11.3"
     misc_feature   1..2252
                     /note="assembly_fragment:00042"
     misc_feature   2353..73816
                     /note="assembly_fragment:04347
                     fragment_chain:1"
     misc_feature   73917..17476
                     /note="assembly_fragment:01075
                     fragment_chain:1"

           1   2252: contig of 2252 bp in length
           2253   2352: gap of 100 bp
           2353 73816: contig of 71464 bp in length
           73817 73916: gap of 100 bp
           73817 17476: gap of 100 bp
           73917 17476: contig of 100560 bp in length

BASE COUNT   53414 a   34556 c   35059 g   51237 t   200 others
ORIGIN

Query Match
Best Local Similarity 30.4%;   Score 29.2;   DB 2;   Length 174476;
Matches 52; Conservative 57.8%; Pred. No. 29; Mismatches 38; Indels 0; Gaps 0;

Qy   2 tttccttcttcgtcgggcgccttggcatggagtggaggaataaagaaggagcgattg 61
          ||||| || |||| |||||| |||||||||| || | | | ||| | ||||
Db 38041 TGTGCANCTATANGTCCTCGGCTTCTCGGACCCTCGAGGTGGCGAGAATCACATGAGCCAAAG 38100

Qy   62 gctgtcgaatggctcagaactgctggagt 91
          ||||| | ||| | | || ||| |||
Db 38101 GAATTCGAGGCTGCATTGAACCACTCCACT 38130

RESULT 15
ACO92000
LOCUS   AC092000   198606 bp   DNA   HTG   27-JUL-2001
DEFINITION Homo sapiens chromosome UNK clone RP11-346I7, WORKING DRAFT
            SEQUENCE, 6 unordered pieces.

ACCESSION   AC092000
VERSION   AC092000.2 GI:15022734
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 198606)
  AUTHORS   Waterston,R.H.
  TITLE   The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 198606)
  AUTHORS   Waterston,R.H.
  TITLE   Direct Submission
  JOURNAL   Submitted (10-JUN-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT   On Jul 27, 2001 this sequence version replaced gi:14334202.

            -------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            ----------------- Project Information -----------------
            Center project name: H_NH0346I07
            ------------------ Summary Statistics -----------------
            Sequencing vector: M13; 6%
            Sequencing vector: plasmid; 94%
            Chemistry: Dye-primer ET; 1% of reads
            Chemistry: Dye-terminator Big Dye; 98% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 195038 bases at least Q40
            Consensus quality: 195597 bases at least Q30
            Consensus quality: 196504 bases at least Q20
            Insert size: 199522; sum-of-contigs
            Insert size: 8988; agarose-fp
            Quality coverage: 9.16 in Q20 bases; agarose-fp
            Quality coverage: 8.54 in Q20 bases; sum-of-contigs

            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.

FEATURES   Location/Qualifiers
     source   1..198606
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="UNK"
                     /clone="RP11-346I7"

            1   1041: contig of 1041 bp in length
            1042   1141: gap of unknown length
            1142   2347: contig of 1206 bp in length
            2348   2447: gap of unknown length
            2448   3591: contig of 1144 bp in length
            3592   3691: gap of unknown length
            3692 21336: contig of 17645 bp in length
            21337 21436: gap of unknown length
            21437 68418: contig of 46982 bp in length
            68419 68518: gap of unknown length
            68519 198606: contig of 130088 bp in length

     misc_feature   1..1041
                     /note="assembly_name:Contig92"
     misc_feature   1142..2347
                     /note="assembly_name:Contig4"
     misc_feature   2448..3591
                     /note="assembly_name:Contig7"
     misc_feature   3692..21336
                     /note="assembly_name:Contig99"

misc_feature   21437..68418
                /note="assembly_name:Contig10"
misc_feature   68519..198606
                /note="assembly_name:Contig11
                clone_end:SP6
                vector_side:right"
BASE COUNT   54258 a   50128 c   46598 t   510 others
ORIGIN

Query Match   30.4%;   Score 29.2;   DB 2;   Length 198606;
Best Local Similarity   62.2%;   Pred. No. 29;
Matches   46;   Conservative   0;   Mismatches   28;   Indels   0;   Gaps   0;

Qy   1 gtttccttcttctgtcgggcgccttggcatggagtggaggaataagaaaggagcgatt 60
          ||| ||| | | | | | | | | ||||||||| ||| || ||
Db 61635 GTTCCCTGCCTCTCCTTGTCCTCACTGGCTTTTGGTGTGAAGAGGAAAAGAAGAAGCCCTT 61694

Qy   61 ggctgtcgatggtg 74
          ||| | | |||||
Db 61695 GGCAGGGCATGGTG 61708

Search completed: February 22, 2002, 08:04:32
Job time: 1642 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2002, 07:37:10 ; Search time 169.15 Seconds
                                              (without alignments)
                                              486.569 Million cell updates/sec

Title: US-09-509-482-1
Perfect score: 96
Sequence: 1 gtttccttcttctgtcgggg.............cagaactgctggagtggagg 96

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Database :   N_Geneseq_1101:*
               1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
               2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
               3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
               4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
               5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
               6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
               7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
               8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
               9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
              10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
              11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
              12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
              13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
              14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
              15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
              16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
              17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
              18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
              19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
              20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
              21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
              22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

SUMMARIES

Result NO.	Score	% Query Match	Length	DB	ID	Description
1	96	100.0	96	20	AAX34787	Human vitamin D re
2	96	100.0	1382	20	AAX34789	Nucleotide sequenc
3	96	100.0	1463	20	AAX34788	Nucleotide sequenc
4	96	100.0	1534	20	AAX34790	Nucleotide sequenc
5	27.4	28.5	1037	21	AAF07559	Fusarium venenatum
6	27.2	28.3	2506	22	AAI60569	Human polynucleoti
7	27.2	28.3	2858	22	AAH18315	Human cDNA sequenc
8	27.2	28.3	2900	22	AAI58783	Human polynucleoti
9	27.2	28.3	3357	21	AAA53921	Human polynucleoti
10	26.8	27.9	461	21	AAA31710	Type IV adenylyl c
11	26.8	27.9	1731	21	AAZ34973	Soybean lycopene c

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

12	26.8	27.9	8082	14	AAQ37740	Complete human c-m
13	26.8	27.9	8082	18	AAT71259	Human c-myb oncoge
14	26.8	27.9	8082	19	AAV30099	Complete nucleotid
15	26.8	27.9	8082	19	AAV20463	Human c-myc oncoge
16	26.8	27.9	8082	21	AAZ60822	Nucleotide sequenc
17	26.6	27.7	1366	20	AAZ27743	Human DNA marker c
18	26.6	27.7	25423	22	AAI57656	Human colorectal c
19	26.6	27.7	25424	22	AAI57657	Human colorectal c
20	26.4	27.5	3423	16	AAQ87798	XAP-1 DNA, part of
21	26.4	27.5	6519	17	AAT37401	Human DOCK180 cDNA
22	26.2	27.3	1065	22	AAH66251	C glutamicum codin
23	26.2	27.3	1438	21	AAZ45336	DNA encoding a man
24	26.2	27.3	1470	21	AAZ45335	DNA encoding a Bac
25	26.2	27.3	1572	22	AAC99932	Human secreted pro
26	26.2	27.3	3102	22	AAH66250	C glutamicum codin
27	26.2	27.3	349980	22	AAH68527	C glutamicum codin
28	26.2	27.3	349980	22	AAH68528	C glutamicum codin
29	26	27.1	756	22	AAF67751	Corynebacterium gl
30	26	27.1	33140	22	AAF26536	Genomic fragment #
31	26	27.1	349940	22	AAH68527	C glutamicum codin
32	25.8	26.9	452	21	AAH50276	Human LSR gene 5'
33	25.8	26.9	3630	17	AAT31289	Bovine poly-immuno
34	25.6	26.7	1032	22	AAF74217	DNA encoding enviro
35	25.6	26.7	2955	21	AAC77554	Human ORFX ORF3109
36	25.6	26.7	6075	21	AAA88866	Human protein tyro
37	25.4	26.5	408	21	AAA31976	Plant microsatelli
38	25.4	26.5	2714	13	AAQ23563	Proctase-A gene.
39	25.4	26.5	22976	20	AAX83426	Genomic region con
40	25.2	26.2	1918	21	AAA08584	Human cytoskeleton
41	25	26.0	1201	22	AAH19242	Human secreted pro
42	25	26.0	1223	22	AAH19194	Human secreted pro
43	24.8	25.8	1132	21	AAC33859	Arabidopsis thalia
44	24.8	25.8	2364	22	AAC66909	Human EXMAD-20 cod
45	24.8	25.8	6703	22	AAF28520	Genomic fragment #

ALIGNMENTS

RESULT 1
AAX34787

ID   AAX34787 standard; DNA; 96 BP.
XX
AC   AAX34787;
XX
DT   06-JUL-1999 (first entry)
XX
DE   Human vitamin D receptor (VDR) gene exon 1d sequence.
XX
KW   Vitamin D receptor; VDR; hVDR; variant; isoform; SRC-1; GRIP-1; TFIIB;
KW   cofactor; human; ss.
XX
OS   Homo sapiens.
XX
PN   WO9916872-A1.
XX
PD   08-APR-1999.
XX
PF   29-SEP-1998; 98WO-AU00817.
XX
PR   29-SEP-1997; 97AU-0009500.
XX
PA   (GARV-) GARVAN INST MEDICAL RES.
XX
PI   Crofts LA, Eisman JA, Hancock MS, Morrison NA;
XX
DR   WPI; 1999-263693/22.
XX
PT   New polynucleotides which encode novel isoforms of the human vitamin
PT   D receptor or variant transcripts for hVDR
XX
PS   Claim 24; Page 33; 56pp; English.

CC   The invention relates to isolated polynucleotides which encode novel
CC   isoforms of the human vitamin D receptor (VDR) or variant transcripts
CC   for hVDR. The polynucleotides are useful in methods for detecting agonist
CC   and /or antagonist compound of a VDR isoform. An increase or decrease in
CC   activity of the receptor may be detected by measuring changes in
CC   interactions with known cofactors (e.g. through use of the dual hybrid system) or
CC   polynucleotides shown in AAX34787, AAX34791 and AAX34792 (corresponding
CC   to hVDR exon sequences) may be useful as probes for the detection of VDR
CC   variant transcripts and for assessing cell or tissue-specific expression
CC   of variant transcripts.
XX
SQ   Sequence 96 BP; 19 A; 16 C; 37 G; 24 T; 0 other;

Query Match   100.0%;   Score 96;   DB 20;   Length 96;
Best Local Similarity   100.0%;   Pred No. 3.5e-24;
Matches   96;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 gtttcctcttctgtcgggggcgccttggccatgggatggaggaggaataagaaaggaggccgatt 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 gtttcctcttctgtcgggggcgccttggccatgggatggaggaggaataagaaaggaggccgatt 60

Qy   61 ggctgtcgatggtgctcaagaactgctggagtggggg 96
             |||||||||||||||||||||||||||||||||||||
Db   61 ggctgtcgatggtgctcaagaactgctggagtggggg 96

RESULT   2

ID   AAX34789 standard; DNA; 1382 BP.
XX
AC   AAX34789;
XX
DT   06-JUL-1999 (first entry)
XX
DE   Nucleotide sequence of human vitamin D receptor (VDR) gene transcript 9.
XX
KW   Vitamin D receptor; VDR; hVDR; variant; isoform; SRC-1; GRIP-1; TFIIB;
KW   cofactor; human; ds.
XX
OS   Homo sapiens.
XX
PN   WO9916872-A1.
XX
PD   08-APR-1999.
XX
PF   29-SEP-1998;   98WO-AU00817.
XX
PR   29-SEP-1997;   97AU-0009500.
XX
PA   (GARV-) GARVAN INST MEDICAL RES.
XX
PI   Crofts LA, Eisman JA, Hancock MS, Morrison NA;
XX
DR   WPI; 1999-263693/22.
DR   P-PSDB; AAY09036.
XX
PT   New polynucleotides which encode novel isoforms of the human vitamin
PT   D receptor or variant transcripts for hVDR
XX
PS   Claim 4; Fig 6; 56pp; English.

The invention relates to isolated polynucleotides which encode novel
isoforms of the human vitamin D receptor (VDR) or variant transcripts
for hVDR. The polynucleotides are useful in methods for detecting agonist
and /or antagonist compound of a VDR isoform. An increase or decrease in
activity of the receptor may be detected by measuring changes in
interactions with known cofactors (e.g. SRC-1, GRIP-1 and TFIIB) or
polynucleotides shown in AAX34787, AAX34791 and AAX34792 (corresponding
to hVDR exon sequences) may be useful as probes for the detection of VDR
variant transcripts and for assessing cell or tissue-specific expression

CC   of variant transcripts. The present sequence represents the nucleotide
CC   sequence of hVDR gene transcript 9.
XX
SQ   Sequence 1382 BP; 310 A; 409 C; 379 G; 284 T; 0 other;

Query Match   100.0%;   Score 96;   DB 20;   Length 1382;
Best Local Similarity   100.0%;   Pred. No. 7.3e-24;
Matches   96;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 gtttcctcttctgtcgggggcgccttggccatgggatggaggaggaataagaaaggaggccgatt 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 gtttcctcttctgtcgggggcgccttggccatgggatggaggaggaataagaaaggaggccgatt 60

Qy   61 ggctgtcgatggtgctcaagaactgctggagtggggg 96
             |||||||||||||||||||||||||||||||||||||
Db   61 ggctgtcgatggtgctcaagaactgctggagtggggg 96

RESULT   3

ID   AAX34788 standard; DNA; 1463 BP.
XX
AC   AAX34788;
XX
DT   06-JUL-1999 (first entry)
XX
KW   Vitamin D receptor; VDR; hVDR; variant; isoform; SRC-1; GRIP-1; TFIIB;
KW   cofactor; human; ds.
XX
OS   Homo sapiens.
XX
PN   WO9916872-A1.
XX
PD   08-APR-1999.
XX
PF   29-SEP-1998;   98WO-AU00817.
XX
PR   29-SEP-1997;   97AU-0009500.
XX
PA   (GARV-) GARVAN INST MEDICAL RES.
XX
PI   Crofts LA, Eisman JA, Hancock MS, Morrison NA;
XX
DR   WPI; 1999-263693/22.
DR   P-PSDB; AAY09035.
XX
PT   New polynucleotides which encode novel isoforms of the human vitamin
PT   D receptor or variant transcripts for hVDR
XX
PS   Claim 4; Fig 5; 56pp; English.

DE   Nucleotide sequence of human vitamin D receptor (VDR) gene transcript 6.

CC   The invention relates to isolated polynucleotides which encode novel
CC   isoforms of the human vitamin D receptor (VDR) or variant transcripts
CC   for hVDR. The polynucleotides are useful in methods for detecting agonist
CC   and/or antagonist compound of a VDR isoform. An increase or decrease in
CC   activity of the receptor may be detected by measuring changes in
CC   interactions with known cofactors (e.g. through use of the dual hybrid system) or
CC   polynucleotides shown in AAX34787, AAX34791 and AAX34792 (corresponding
CC   to hVDR exon sequences) may be useful as probes for the detection of VDR
CC   variant transcripts. The present sequence represents the nucleotide
CC   sequence of hVDR gene transcript 6.
XX
SQ   Sequence 1463 BP; 325 A; 437 C; 398 G; 303 T; 0 other;

Query Match   100.0%;   Score 96;   DB 20;   Length 1463;
Best Local Similarity   100.0%;   Pred. No. 7.4e-24;
Matches   96;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 gttttccttcttctgtctcgggggccgcctggcagtcggagtggagggaataagaaaagagagccgatt   60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 gttttccttcttctgtctcgggggccgcctggcagtcggagtggagggaataagaaaagagagccgatt   60

Qy   61 ggctgtcgatggtgtgctcagaactgctggagtggagg   96
         ||||||||||||||||||||||||||||||||||||||
Db   61 ggctgtcgatggtgtgctcagaactgctggagtggagg   96

RESULT   4
AAX34790
ID   AAX34790 standard; DNA; 1534 BP.
XX
AC   AAX34790;
XX
DT   06-JUL-1999   (first entry)
XX
DE   Nucleotide sequence of human vitamin D receptor (VDR) gene transcript 10.
XX
KW   Vitamin D receptor; VDR; hVDR; variant; isoform; SRC-1; GRIP-1; TFIIB;
KW   cofactor; human; ds.
XX
OS   Homo sapiens.
XX
PN   WO9916872-A1.
XX
PD   08-APR-1999.
XX
PF   29-SEP-1998;   98WO-AU00817.
XX
PR   29-SEP-1997;   97AU-0009500.
XX
PA   (GARV-) GARVAN INST MEDICAL. RES.
XX
PI   Crofts LA, Eisman JA, Hancock MS, Morrison NA;
XX
DR   WPI; 1999-263693/22.
DR   P-PSDB; AAX09037.
XX
PT   New polynucleotides which encode novel isoforms of the human vitamin
PT   D receptor or variant transcripts for hVDR
XX
PS   Claim 4; Fig 7; 56pp; English.
XX
CC   The invention relates to isolated polynucleotides which encode novel
CC   isoforms of the human vitamin D receptor (VDR) or variant transcripts
CC   for hVDR. The polynucleotides are useful in methods for detecting agonist
CC   and/or antagonist compound of a VDR isoform. An increase or decrease in
CC   activity of the receptor may be detected by measuring changes in
CC   interactions with known cofactors (e.g. SRC-1, GRIP-1 and TFIIB) or
CC   unknown cofactors (e.g. through use of the dual hybrid system). The
CC   polynucleotides shown in AAX34787, AAX34791 and AAX34792 (corresponding
CC   to hVDR exon sequences) may be useful as probes for the detection of VDR
CC   variant transcripts and for assessing cell or tissue-specific expression
CC   of variant transcripts. The present sequence represents the nucleotide
CC   sequence of hVDR gene transcript 10.
XX
SQ   Sequence 1534 BP; 345 A; 452 C; 422 G; 315 T; 0 other;

Query Match   100.0%;   Score 96;   DB 20;   Length 1534;
Best Local Similarity   100.0%;   Pred. No. 7.5e-24;
Matches   96;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 gtttccttcttctgtctcgggggccgcctggcagtcggagtggagggaataagaaaagagagccgatt   60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 gtttccttcttctgtctcgggggccgcctggcagtcggagtggagggaataagaaaagagagccgatt   60

Qy   61 ggctgtcgatggtgtgctcagaactgctggagtggagg   96
         ||||||||||||||||||||||||||||||||||||||
Db   61 ggctgtcgatggtgtgctcagaactgctggagtggagg   96

RESULT   5
AAF07559/c
ID   AAF07559 standard; cDNA; 1037 BP.
XX
AC   AAF07559;
XX
DT   13-MAR-2001   (first entry)
XX
DE   Fusarium venenatum EST SEQ ID NO:82.
XX
KW   Multiple gene expression; filamentous fungal cell; EST;
KW   expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW   Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW   culture condition; environmental stress; spore morphogenesis;
KW   metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS   Fusarium venenatum.
XX
PN   WO200056762-A2.
XX
PD   28-SEP-2000.
XX
PF   22-MAR-2000;   2000WO-US07781.
XX
PR   22-MAR-1999;   99US-0273623.
XX
PA   (NOVO) NOVO NORDISK BIOTECH INC.
PA   (NOVO) NOVO NORDISK AS.
XX
PI   Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR   WPI; 2000-594572/56.
XX
PT   Monitoring differential expression of genes in filamentous fungal cells
PT   uses fluorescence-labeled nucleic acids isolated from the cells and a
PT   substrate of expressed sequence tags -
XX
PS   Claim 86; Page 405; 3161pp; English.
XX
CC   The present invention describes a method for monitoring differential
CC   expression of genes in a first filamentous fungal (FF) cell relative to
CC   expression of the same genes in one or more second filamentous fungal
CC   cells. The method uses fluorescence-labeled nucleic acids isolated from
CC   the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC   are used in the methods for monitoring differential expression of genes
CC   in a first filamentous fungal (FF) cell relative to expression of the
CC   same genes in one or more second filamentous fungal cells. Monitoring
CC   the global expression of genes from FF cells allows the production
CC   potential of the microorganisms to be improved. New genes may be
CC   discovered, possible functions of unknown open reading frames can be
CC   identified and gene copy number variation and stability can be
CC   monitored. The expression of genes can be used to study how FF cells
CC   adapt to changes in culture conditions, environmental stress, spore
CC   morphogenesis, recombination, metabolic or catabolic pathway
CC   engineering. Using ESTs provides several advantages over genomic or
CC   random cDNA clones including elimination of redundancy as one spot on an
CC   array equals one gene or open reading frame, and organisation of the
CC   microarrays based on function of the gene products to facilitate
CC   analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC   Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC   niger; AAF11854 to AAF11878 represents ESTs from Aspergillus oryzae; and
CC   AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC   all specifically claimed in the present invention.
XX
SQ   Sequence 1037 BP; 260 A; 296 C; 248 G; 223 T; 10 other;

Query Match   28.5%;   Score 27.4;   DB 21;   Length 1037;
Best Local Similarity   69.8%;   Pred. No. 4.1;
Matches   37;   Conservative   0;   Mismatches   16;   Indels   0;   Gaps   0;

Qy   25 ttgcatggagtggaggaggaataagaaaagagagcgattggctgtcgatggtgctc   77

Left column

```
         Db    180  TTGGCATAGAGAGGAATATCGTGGAAAGAGAGACGGGCCTGTCCATCCTTCTC  128
                    ||||||| ||| ||    |||||| || ||||| | ||||||| || ||||
```

Query Match　　　　28.38;　　　Score 27.2;　　DB 22;　　Length 2506;

SQ   Sequence 2506 BP; 391 A; 670 C; 802 G; 643 T; 0 other;

XX
CC   specification. The sequence data for this patent did not form part of the printed
CC   Note: The sequence data for this patent did not form part of the printed
CC   C.N.S disorders.
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, drug screening,
CC   and thrombolytic activity, cancer diagnosis and therapy, haemostatic
CC   utilisation of the activities such as: Immune system suppression, where the
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   localised neuropathies and central nervous system diseases, such as
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   The invention relates to human nucleic acids and polypeptides (AAI57798-AAI61369) and

PT   Claim 1; SEQ ID NO 4558; 10078pp; English.

XX
PS   (HYSE-) HYSEQ INC.

PT   such as central nervous system injuries -
PT   Novel nucleic acids and polypeptides, useful for treating disorders

XX
DR   P-PSDB; AAM41413.
DR   WPI; 2001-442253/47.

PR   29-NOV-2000; 2000US-0727344.
PR   19-OCT-2000; 2000US-0693036.
PR   14-SEP-2000; 2000US-0662191.
PR   03-AUG-2000; 2000US-0653450.
PR   19-JUL-2000; 2000US-0620312.
PR   09-JUL-2000; 2000US-0598042.
PR   25-APR-2000; 2000US-0552317.
PR   21-JAN-2000; 2000US-0488725.

XX
PF   26-DEC-2000; 2000WO-US34263.

XX
PD   26-JUL-2001.

XX
PN   WO200153312-A1.

XX
OS   Homo sapiens.

XX
DE   Human polynucleotide SEQ ID NO 4558.

XX
DT   22-OCT-2001 (first entry)

XX
AC   AAI60569;

XX
ID   AAI60569 standard; cDNA; 2506 BP.

XX
RESULT   6
AA160569

PI   Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI   Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI   Zhao QA, Zhou P, Goodrich R, Drmanac RT;

KW   leukaemia; ss.
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
```

## Right column

Best Local Similarity 58.8%; Conservative 0; Pred. No. 6.2; Indels 0; Gaps 0;
Matches 47; Conservative 0; Mismatches 33;

```
Qy   17  gggggcgccttggccatggagttggaggaataagaaaagaggagcgattggcttgtcgatggtggtgct  76
             ||||||||||||||||||||||| ||||| |||||| |||  |||| |||||||||| ||||||
Db   1032  ggagcagctggccttgcagatgtagcagaagaagaactggaggcctgcttagaggggagtgct  1091

Qy   77  cagaactgctggagtggagg  96
            |||||||||||||||||||
Db   1092  ggtagccgcaggggtggagg  1111
```

RESULT   7
AAH18315/c

ID   AAH18315 standard; cDNA; 2858 BP.

XX
AC   AAH18315;

XX
DT   26-JUN-2001 (first entry)

XX
DE   Human cDNA sequence SEQ ID NO:18316.

XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX
OS   Homo sapiens.

XX
PN   EP1074617-A2.

XX
PD   07-FEB-2001.

XX
PF   28-JUL-2000; 2000EP-0116126.

PR   29-JUL-1999; 99JP-0248036.
PR   27-AUG-1999; 99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.

XX
DR   WPI; 2001-318749/34.

XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX
PS   (HELI-) HELIX RES INST.

XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -

XX
PT   Claim 8; SEQ ID NO 18316; 2537pp + CD ROM; English.

XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides in the specification;
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers allow obtaining of the full-length
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers are also useful for the
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification

CC   of the present invention.
XX
SQ   Sequence 2858 BP; 695 A; 948 C; 760 G; 455 T; 0 other;

Query Match   28.3%; Score 27.2; DB 22; Length 2858;
Best Local Similarity   58.8%; Pred. No. 6.5;
Matches 47; Conservative   0; Mismatches 33; Indels   0; Gaps   0;

Qy   17 ggggcgccttggcatggagtggaggaataagaaaggaggacgagttggcgtcgatggtgct 76
          ||| || | ||| |||| || || ||||||||||| ||||| | ||||||||||||
Db   1844 GGAGCAGCTGGCCTTGCCAGATGTAGGCAGAGAAGAACTGGAGGCGCCTGCTTAGAGGGAGTGCT 1785

Qy   77 cagaactgctggagtggagg 96
          || ||| || |||||||||
Db   1784 GGTAGCCGCAGGGGTGGAGG 1765

RESULT   8
AAI58783/c

ID   AAI58783 standard; cDNA; 2900 BP.
XX
AC   AAI58783;
XX
DT   22-OCT-2001 (first entry)
XX
DE   Human polynucleotide SEQ ID NO 986.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PR   26-DEC-2000; 200WO-US34263.
PR   29-NOV-2000; 200US-0727344.
PR   19-OCT-2000; 200US-0693036.
PR   14-SEP-2000; 200US-0661191.
PR   03-AUG-2000; 200US-0653450.
PR   19-JUL-2000; 200US-0620312.
PR   09-JUL-2000; 200US-0598042.
PR   25-APR-2000; 200US-0552337.
PR   21-JAN-2000; 200US-0488725.
XX
DR   WPI; 2001-442255/47.
DR   P-PSDB; AAM39627.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI   Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI   Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
PS   Claim 1; SEQ ID NO 986; 1007bp; English.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence 2900 BP; 725 A; 945 C; 780 G; 450 T; 0 other;

Query Match   28.3%; Score 27.2; DB 22; Length 2900;
Best Local Similarity   58.8%; Pred. No. 6.5;
Matches 47; Conservative   0; Mismatches 33; Indels   0; Gaps   0;

Qy   17 ggggcgccttggcatggagtggaggaataagaaaggaggacgagttggcgtcgatggtgct 76
          ||| || | ||| |||| || || ||||||||||| ||||| | ||||||||||||
Db   1874 GGAGCAGCTGGCCTTGCCAGATGTAGGCAGAGAAGAACTGGAGGCGCCTGCTTAGAGGGAGTGCT 1815

Qy   77 cagaactgctggagtggagg 96
          || ||| || |||||||||
Db   1814 GGTAGCCGCAGGGGTGGAGG 1795

RESULT   9
AAA53921/c

ID   AAA53921 standard; cDNA; 3357 BP.
XX
AC   AAA53921;
XX
DT   03-JAN-2001 (first entry)
XX
DE   Type IV adenylyl cyclase coding sequence.
XX
KW   Adenylyl cyclase; type I; type II; recombinant; enzyme; cAMP;
KW   cyclic AMP; adenosine monophosphate; screening; stimulation;
KW   inhibition; treatment; cholera; pituitary tumour; heart failure;
KW   ischaemia; endocrine disorder; cell necrosis;
KW   pseudohypoparathyroidism; endocrine deficiency; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   110..3304
FT   /*tag= a
FT   /product= Type IV adenylyl cyclase
XX
PN   US6107076-A.
XX
PD   22-AUG-2000.
XX
PR   04-OCT-1996; 96US-0726214.
PR   04-OCT-1995; 95US-0005498.
XX
PA   (TEXA) UNTV TEXAS SYSTEM.
XX
PI   Gilman AG, Tang W;
XX
DR   WPI; 2000-578539/54.
DR   P-PSDB; AAB02008.
XX
PS   Disclosure; Columns 57-60; 73pp; English.
XX
PT   Novel soluble mammalian polypeptide composition comprising adenylyl
PT   cyclase activity for screening stimulators and inhibitors of adenylyl
PT   cyclase, is activated by Gsalpha
XX
CC   A recombinant Adenylyl cyclase is described which lacks membrane
CC   bound domains. Separation and purification of the recombinant
CC   enzyme is much easier compared with wild type enzymes and the

CC   recombinant enzyme is more stable than the wild type enzyme which
CC   allows easier screening of compounds that stimulate and inhibit
CC   Adenylyl cyclase activity. The recombinant adenylyl cyclase comprises
CC   a chimera of adenylyl cyclase C_1 and C_2 domains linked covalently.
CC   The domains may be linked by a linker peptide. The recombinant
CC   adenylyl cyclase is useful for screening inhibitors and stimulators
CC   of adenylyl cyclase activity. Inhibitors of the enzyme are useful for
CC   treating cholera, pituitary tumors, heart failure, ischaemia,
CC   endocrine disorders and cell necrosis. Stimulators of adenylyl
CC   cyclase are useful for treating pseudohypoparathyroidism and other
CC   endocrine deficiencies.
XX
SQ   Sequence 3357 BP; 707 A; 985 C; 904 G; 761 T; 0 other;

Query Match   28.3%;   Score 27.2;   DB 21;   Length 3357;
Best Local Similarity   64.1%;   Pred. No. 6.8;
Matches   41;   Conservative   0;   Mismatches   23;   Indels   0;   Gaps   0;

Qy   30 atggagtggaggaattaagagaaaggagacgagattggtcgatggtgtctcagaactgctgga 89
Db 1983 AAGAGGAGGAGGAAGAGAAGAGAAGGAAGGTCATGCGTCGTAGGTGGTGGCCAGAGCTGGGGGC 1924

Qy   90 gtgg 93
Db 1923 CTGG 1920

RESULT 10
ID   AAA31710/c
     AAA31710 standard; DNA; 461 BP.
AC   AAA31710;
XX
DT   05-JUL-2000 (first entry)
XX
DE   Plant microsatellite marker #671.
XX
KW   Plant microsatellite sequence; core repeat sequence; detection; probe;
KW   DNA polymorphism; genome mapping; physical mapping; fingerprinting;
KW   variety identification; genetic variability evaluation; primer; ss.
XX
OS   Eucalyptus grandis.
XX
PN   WO9967421-A1.
XX
PD   29-DEC-1999.
XX
PF   25-JUN-1999; 99WO-NZ00092.
XX
PR   25-JUN-1998; 98US-0105307.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Havukkala IJ, Bloksberg LN, Glenn M;
XX
DR   WPI; 2000-116958/10.
XX
PT   New plant microsatellite markers and associated flanking species for
PT   the detection of polymorphic genetic markers -
XX
PS   Claim 1; Page 274; 392pp; English.
XX
CC   Sequences AAA31040-A32093 represent novel plant microsatellite sequences
CC   and associated flanking species. The sequences comprise a central core
CC   repeat sequence, especially selected from the sequences AAA32094-A32096
CC   with left and right flanking sequences. The polynucleotide sequences
CC   can be used in the detection of DNA polymorphisms, in genome mapping,
CC   in physical mapping, in positional cloning of genes, in variety
CC   identification and in evaluation of genetic variability within and
CC   between plant tissues, populations, cultivars, species and species
CC   groups. They may also be used to design hybridization probes for

CC   oligonucleotide fingerprinting and library screening and to design
CC   primers for microsatellite-primed PCR. Microsatellite markers are
CC   useful to locate specific economically useful genes in plant genomes.
XX
SQ   Sequence 461 BP; 92 A; 117 C; 144 G; 108 T; 0 other;

Query Match   27.9%;   Score 26.8;   DB 21;   Length 461;
Best Local Similarity   55.3%;   Pred. No. 5.3;
Matches   52;   Conservative   0;   Mismatches   42;   Indels   0;   Gaps   0;

Qy   3 ttccttcttctgtcggggcggccttgggcatggagtggaggaggaattaagaaaggagacgattgg 62
Db   300 TTTCTCTCTCTCTCTCTCTCCGCCTTCGAATCGCGTTGGACGAAAAAAAGAAAAACAACTCGAGA 241

Qy   63 ctgtcgatggtgtcaagaactgctggggtgggg 96
Db   240 GGGCGCGCCCCAACTTAGACCAGCAGCAGCAGCGG 207

RESULT 11
ID   AAZ34973/c
     AAZ34973 standard; cDNA; 1731 BP.
AC   AAZ34973;
XX
DT   28-FEB-2000 (first entry)
XX
DE   Soybean lycopene cyclase cDNA.
XX
KW   Lycopene cyclase; soybean; transgenic plant; carotenoid;
KW   beta-carotene; pigment; ss.
XX
OS   Glycine max.
XX
FH   Key   Location/Qualifiers
FT   CDS   89..1570
FT   /*tag= a
XX
PN   WO9955887-A2.
XX
PD   04-NOV-1999.
XX
PF   16-APR-1999; 99WO-US08384.
XX
PR   24-APR-1998; 98US-0083042.
XX
PA   (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI   Cahoon RE, Kinney AJ, Pearlstein RW, Williams ME;
XX
DR   WPI; 2000-062037/05.
DR   P-PSDB; AAY32325.
XX
PT   Novel carotenoid biosynthesis enzyme polynucleotides and polypeptides
PT   used to identify inhibitors and to modulate expression of the enzyme -
XX
PS   Claim 12; Page 49-50; 61pp; English.
XX
CC   This is the nucleotide sequence of the cDNA insert of clone
CC   sfl1.pk0034.c1 encoding an entire soybean lycopene cyclase (see
CC   AAY32325). The clone was isolated from an immature flower cDNA
CC   library. The invention provides novel carotenoid biosynthesis
CC   enzymes, specifically beta-carotene hydroxylase, lycopene cyclase
CC   and lycopene epsilon cyclase (see AAY32318-26 and AAY32286-88), and
CC   polynucleotides encoding them (see AAZ34967-74). The enzymes may be
CC   prepared recombinantly and used to raise antibodies, for use in
CC   detecting the enzymes. The polynucleotides may be used to create
CC   transgenic plants in which the beta-carotene hydroxylase, lycopene
CC   cyclase and lycopene epsilon cyclase levels are present at higher or
CC   lower levels than normal, or in cell types or developmental processes
CC   where they are not normally found. This would alter the level of
CC   alpha-carotene or beta-carotene in those cells. The enzymes can

CC   also be used as targets to facilitate the design and/or
CC   identification of inhibitors of those enzymes that may be useful
CC   as herbicides. This is desirable because inhibition of any of the
CC   enzymes could lead to an inhibition of plant growth. The
CC   polynucleotides also serve as a source of probes and primers, which
CC   are useful for genetic mapping, and to isolate homologous sequences
CC   from other species.
XX
SQ   Sequence 1731 BP; 446 A; 325 C; 429 G; 530 T; 1 other;

Query Match   27.9%; Score 26.8; DB 21; Length 1731;
Best Local Similarity   61.4%; Pred. No. 7.7;
Matches 43; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy   5 cctcttctgtcggggcgcctggccatggagtcggaggaaataagaaaaggaggcgattggct 64
           ||||||||| ||| |||||||| |||||||| ||| ||| || |||| |||| ||||||
Db   206 CCCTCTTGTTCTTCTTGAATGCTGTGTGAGGGGTTTGAGGAAGAGAGGGAAATGGAGTGGGTGCAGGG 147

Qy   65 gtcgatgggg 74
           || |||||
Db   146 GTACTTGGTG 137

RESULT 12
AAQ37740
ID   AAQ37740 standard; DNA; 8082 BP.
XX
AC   AAQ37740;
XX
DT   30-JUN-1993 (first entry)
XX
DE   Complete human c-myc DNA sequence.
XX
KW   Cellular division cycle; cdc; antisense oligonucleotide; ss.
XX
OS   Homo sapiens.
XX
FH
FT   Key   Location/Qualifiers
FT
FT   misc_feature   6002..6004
FT   /tag= a
FT   /note= "unclear from sequence given in specification"
XX
XX
PN   US7821415-A.
XX
PD   01-JAN-1993.
XX
PF   14-JAN-1992; 92US-0821415.
XX
PR   14-JAN-1992; 92US-0821415.
XX
PA   (USSH) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Epstein S, Speir E, Unger E;
XX
PT   Inhibition of re-stenosis of blood vessel - after mechanical
PT   treatment, to reduce stenosis, using anti-sense oligo:nucleotide(s)
XX
PS   Disclosure; Page 22; 41pp; English.
XX
CC   This the complete DNA sequence of human c-myc which may be used to
CC   derive potential target antisense oligonucleotide sequences for
CC   inhibiting translation of cellular division cycle (cdc) gene
CC   products. These may be used in a method of inhibiting restenosis
CC   of a mammalian blood vessel after mechanical treatment to reduce
CC   a stenosis, e.g. coronary balloon angioplasty.
XX
SQ   Sequence 8082 BP; 1849 A; 2116 C; 2133 G; 1982 T; 2 other;

Query Match   27.9%; Score 26.8; DB 14; Length 8082;

Best Local Similarity   61.4%; Pred. No. 12;
Matches 43; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy   17 ggggcgccttggccatggagtcggaggaaataagaaaaggaggcgattggctgcgatggtgct 76.
           |||||| ||| |||||||| |||||||| ||| ||| || |||| |||| ||||||
Db   5596 GGGGCGGGGTGGCAGGGAGTGTATGGAGGATGAAGGGGATTGAGAGGAGGAGGATTGATCTGAGAGAGTGA 5655

Qy   77 cagaactgct 86
           ||| |||| |||
Db   5656 atgaattgct 5665

RESULT 13
AAT71259
ID   AAT71259 standard; DNA; 8082 BP.
XX
AC   AAT71259;
XX
PN   WO9725860-A1.
XX
DT   30-MAR-1998 (first entry)
XX
DE   Human c-myb oncogene.
XX
KW   Cognate transgene; human; c-myb; lymphoma; cellular immunogen; cancer;
KW   self-determinant immunoreactivity; cancer vaccination; breast carcinoma;
KW   colon carcinoma; immunotherapy; proto-oncogene; ss.
XX
OS   Homo sapiens.
XX
DT   24-JUL-1997.
XX
PD   13-JAN-1997; 97WO-US00582.
XX
PF   13-JAN-1997; 97WO-US00582.
XX
PR   19-JAN-1996; 96US-0010262.
XX
DR   WPI; 1997-384993/35.
XX
PA   (UYAL-) UNIV ALLEGHENY HEALTH SCI.
XX
PI   England JM, Halpern MS;
XX
PT   Proto-oncogene immunogen - used in vaccine for the prevention and
PT   treatment of cancer
XX
PS   Disclosure; Page 62-66; 81pp; English.
XX
CC   This sequence represents the human c-myb cognate transgene (CTG).
CC   Deletion of amino acids 129-144 of the encoded protein renders the CTG
CC   non-transforming. This sequence can be used in the cellular immunogen of
CC   the invention. The cellular immunogen of the invention is for immunising
CC   against the product of a target proto-oncogene, over-expression of which
CC   is associated with cancer, comprises host cells transfected with a
CC   construct containing at least one transgene related to the proto-oncogene
CC   and driven by a strong promoter. The product of the transgene induces
CC   immunoreactivity to host self-determinants on the product of
CC   proto-oncogene. The cellular immunogens are used for protective
CC   vaccination against cancer (e.g. carcinoma of breast or colon, or various
CC   lymphomas) and for immunotherapy of cancer. Use of the immunogen
CC   eliminates the need to isolate immunogenic, HLA host-matched peptides.
CC   The method is not based on immune recognition of a determinant defined by
CC   a cancer-specific mutation and generates a systemic (anti-metastatic)
CC   response.
XX
SQ   Sequence 8082 BP; 1850 A; 2115 C; 2135 G; 1982 T; 0 other;

Query Match   27.9%; Score 26.8; DB 18; Length 8082;
Best Local Similarity   61.4%; Pred. No. 12;
Matches 43; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy   17 ggggcgccttggccatggagtcggaggaaataagaaaaggaggcgattggctgcgatggtgct 76
           |||||| ||| |||||||| |||||||| ||| ||| || |||| |||| ||||||

```
Db  5596 gggcgcgggtggcagggagtgtatgaatgaggaggaggattgatctctgagagtga 5655

Qy    77 cagaactgct 86
         ||| ||||
Db  5656 atgaattgct 5665

RESULT 14
AAV30099
ID   AAV30099 standard; DNA; 8082 BP.
XX
AC   AAV30099;
XX
DT   11-AUG-1998 (first entry)
XX
DE   Complete nucleotide sequence of the human c-myc gene.
XX
KW   Human; c-myc gene; inhibition; growth; smooth muscle cell; migration;
KW   restenosis; blood vessel; reduction; stenosis;
KW   antisense oligonucleotide; cardiac angioplasty; ds.
XX
OS   Homo sapiens.
XX
PN   US5756476-A.
XX
PD   26-MAY-1998.
XX
PF   14-JAN-1994; 94US-0187785.
XX
PR   26-JAN-1994; 94US-0187785.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Epstein SE, Speir EH, Unger EF;
XX
DR   WPI; 1998-321578/28.
XX
PT   Inhibition of smooth-muscle cell proliferation, especially
PT   restenosis - comprises contacting cells with antisense
PT   oligonucleotides of c-myc gene product.
XX
PS   Disclosure; Columns 15-22; 23pp; English.
XX
CC   The present sequence represents the complete nucleotide sequence of the
CC   human c-myc gene. The specification describes methods for inhibiting
CC   the growth of a human smooth muscle cell, inhibiting the migration of
CC   a human smooth muscle cell and inhibiting restenosis of a blood vessel
CC   in a human after mechanical treatment to the vessel to reduce stenosis.
CC   The methods comprise contacting the cell or vessel with a synthetic
CC   antisense oligonucleotide directed against a c-myc gene product.
CC   The antisense oligonucleotides are used for inhibiting restenosis after
CC   cardiac angioplasty.
XX
SQ   Sequence 8082 BP; 1850 A; 2115 C; 2135 G; 1982 T; 0 other;

Query Match        27.9%;   Score 26.8;   DB 19;   Length 8082;
Best Local Similarity   61.4%;   Pred. No. 12;
Matches  43;  Conservative  0;  Mismatches  27;  Indels  0;  Gaps  0;

Qy    17 gggcgccttggcatggagtggagaataagaaaagggagcgattggctgtcgatggtgct 76
         |||||||      |||||||        |||  ||| ||||  |||| |
Db  5596 gggcgcgggtggcagggagtgtatgaatgaggaggaggattgatctctgagagtga 5655

Qy    77 cagaactgct 86
         ||| ||||
Db  5656 atgaattgct 5665

RESULT 15
AAV20463
ID   AAV20463 standard; DNA; 8082 BP.
XX
AC   AAV20463;
XX
DT   17-JUN-1998 (first entry)
XX
DE   Human c-myc oncogene.
XX
KW   Human; oncogene; proto-oncogene; neoplastic disease; anticancer;
KW   cancer; antisense oligonucleotide; c-myc; ds.
XX
OS   Homo sapiens.
XX
PN   US5734039-A.
XX
PD   31-MAR-1998.
XX
PF   15-SEP-1994; 94US-0306691.
XX
PR   15-SEP-1994; 94US-0306691.
XX
PA   (UUJE-) UNIV JEFFERSON THOMAS.
XX
PI   Calabretta B, Skorski T;
XX
DR   WPI; 1998-229882/20.
XX
PT   Anticancer composition comprising two anti-sense oligo:nucleotide(s)
PT   - targetting cytoplasmic and nuclear oncogene(s)
XX
PS   Claim 1; Column 113-120; 92pp; English.
XX
CC   The present sequence represents an oncogene from the present invention.
CC   The present invention describes a composition which comprises two
CC   antisense oligonucleotides. The first oligonucleotide is specific for a
CC   cytoplasmic oncogene or proto-oncogene selected from ras, raf, EGF-1,
CC   c-fms, c-ros, c-kit, c-met, c-trk, c-src, c-abl, bcr-abl, c-fgr and
CC   c-yes. The second oligonucleotide is specific for a nuclear oncogene or
CC   proto-oncogene selected from myc, jun, c-ets, c-fos, c-myb, B-myb,
CC   c-rel, c-vav, c-ski, c-spi, cyclin D1, PML/RAR alpha , AML1/MTG8,
CC   E2A/pri and ALL-1/AF-4. The composition is used for treating cancer.
CC   The combination of antisense oligonucleotides has synergistically
CC   enhanced ability to inhibit growth of cancer cells.
XX
SQ   Sequence 8082 BP; 1850 A; 2115 C; 2135 G; 1982 T; 0 other;

Query Match        27.9%;   Score 26.8;   DB 19;   Length 8082;
Best Local Similarity   61.4%;   Pred. No. 12;
Matches  43;  Conservative  0;  Mismatches  27;  Indels  0;  Gaps  0;

Qy    17 gggcgccttggcatggagtggagaataagaaaagggagcgattggctgtcgatggtgct 76
         |||||||      |||||||        |||  ||| ||||  |||| |
Db  5596 gggcgcgggtggcagggagtgtatgaatgaggaggaggattgatctctgagagtga 5655

Qy    77 cagaactgct 86
         ||| ||||
Db  5656 atgaattgct 5665

Search completed: February 22, 2002, 08:32:05
Job time: 3295 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2002, 07:37:10 ; Search time 85.32 Seconds
                                              (without alignments)
                                              254.828 Million cell updates/sec

Title: US-09-509-482-1
Perfect score: 96
Sequence: 1 gttccttcctctgtccgggg..........cagaactgctggagtggagg 96

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_NA:*
                1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
c 1	27.2	28.3	3357 3	US-08-726-214-7	Sequence 7, Appli
c 2	26.8	27.9	8056 3	US-09-136-605-14	Sequence 14, Appl
c 3	26.8	27.9	8082 3	US-08-306-691B-41	Sequence 41, Appl
c 4	26.8	27.9	8082 1	US-08-187-785-1	Sequence 1, Appl
c 5	26.8	27.9	8082 5	PCT-US93-06251-28	Sequence 28, Appl
c 6	26.6	27.7	1366 4	US-09-018-584A-43	Sequence 43, Appl
c 7	26.4	27.5	6519 1	US-08-588-985-1	Sequence 1, Appl
c 8	26.4	27.5	6519 1	US-08-971-988-1	Sequence 1, Appl
c 9	25.8	26.9	3630 1	US-08-434-000A-5	Sequence 5, Appli
c 10	25.8	26.9	3630 4	US-09-312-157-5	Sequence 5, Appli
c 11	25.8	26.9	11558 5	PCT-US93-06251-23	Sequence 23, Appl
c 12	24.2	25.2	1205 1	US-08-518-878B-36	Sequence 36, Appl
c 13	24.2	25.2	1205 2	US-08-294-522B-37	Sequence 37, Appl
c 14	24.2	25.2	1205 2	US-08-807-861A-36	Sequence 36, Appl
c 15	24.2	25.2	1205 3	US-08-470-868A-36	Sequence 36, Appl
c 16	24.2	25.2	1205 3	US-09-210-681-36	Sequence 36, Appl
c 17	24.2	25.2	1205 3	US-08-946-719A-36	Sequence 36, Appl
c 18	24.2	25.2	24979 2	US-08-147-777-3	Sequence 3, Appli
c 19	24.2	25.2	24979 3	US-08-452-872-3	Sequence 3, Appli
c 20	24.2	25.2	24979 5	PCT-US93-03985-3	Sequence 3, Appli
c 21	24	25.0	3188 4	US-08-943-731-183	Sequence 183, App
c 22	24	25.0	20084 4	US-08-943-731-5	Sequence 5, Appli
c 23	23.8	24.8	1997 2	US-08-969-630-1	Sequence 1, Appli
c 24	23.4	24.4	1412 1	US-08-097-831-1	Sequence 1, Appli
c 25	23.2	24.2	2102 1	US-08-784-651-5	Sequence 5, Appli
c 26	23.2	24.2	3011 1	US-07-821-716-1	Sequence 1, Appli
c 27	23	24.0	1044 4	US-09-094-557-4	Sequence 4, Appli
c 28	22.8	23.8	248 3	US-09-008-979A-4	Sequence 4, Appli
c 29	22.8	23.8	248 4	US-09-460-618-4	Sequence 4, Appli
c 30	22.8	23.8	729 4	US-08-977-865-3	Sequence 3, Appli
c 31	22.8	23.8	1678 4	US-08-650-766-2	Sequence 2, Appli
c 32	22.8	23.8	1954 3	US-08-922-635-2	Sequence 2, Appli
c 33	22.8	23.8	2456 3	US-08-813-150-5	Sequence 5, Appl
c 34	22.8	23.8	2557 4	US-08-464-954A-1	Sequence 1, Appl
c 35	22.8	23.8	3318 3	US-08-650-766-3	Sequence 3, Appli
c 36	22.8	23.8	3318 3	US-08-922-635-3	Sequence 3, Appli
c 37	22.8	23.8	3385 3	US-08-650-766-1	Sequence 1, Appli
c 38	22.8	23.8	3385 3	US-08-922-635-1	Sequence 1, Appli
c 39	22.8	23.8	15202 3	US-09-046-736-3	Sequence 3, Appli
c 40	22.8	23.5	1502 3	PCT-US95-12357A-3	Sequence 3, Appl
c 41	22.6	23.3	1027 5	US-08-002-202-12	Sequence 12, Appli
c 42	22.4	23.3	1356 1	US-08-002-202-16	Sequence 16, Appl
c 43	22.4	23.3	1356 1	US-08-002-202-18	Sequence 18, Appl
c 44	22.4	23.3	1356 1	US-08-002-202-18	Sequence 18, Appl
c 45	22.4	23.3	1356 3	US-08-481-534-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-08-726-214-7/c
Sequence 7, Application US/08726214
Patent No. 6107076
; GENERAL INFORMATION:
; APPLICANT: Tang, Wei-Jen
; APPLICANT: Gilman, Alfred G.
; TITLE OF INVENTION: SOLUBLE MAMMALIAN ADENYLYL CYCLASE
; TITLE OF INVENTION: AND USES THEREFOR
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Arnold, White & Durkee
;   STREET: P.O. Box 4433
;   CITY: Houston
;   STATE: Texas
;   COUNTRY: United States of America
;   ZIP: 77210
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/726,214
;   FILING DATE: Concurrently Herewith
;   CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 60/005,498
;   FILING DATE: 04-OCT-1995
; ATTORNEY/AGENT INFORMATION:
;   NAME: Highlander, Steven L.
;   REGISTRATION NUMBER: 37,642
;   REFERENCE/DOCKET NUMBER: UTSD:450
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (512) 418-3000
;   TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 7:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 3357 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;
US-08-726-214-7

Query Match   28.3%; Score 27.2; DB 3; Length 3357;
Best Local Similarity   64.1%; Pred. No. 1.3;
Matches   41; Conservative   0; Mismatches   23; Indels   0; Gaps   0;

Qy   30 atggagtgggagaataagaaaaaggagcgattggtcgtcgatggtgctcagaactgctgga 89

```
Db  1983  AAGAGGAGGAGGAAGGAAGGAAGGAAGGAAGTGATGCTGTAAGTGGTGGCCAGAGCTGGGGGC  1924
          ||||  ||||| ||||  |||||  |||||         ||||| ||||| |||| |||   ||
Qy    90  gtgg  93
          |||
Db  1923  CTGG  1920
```

RESULT 2
US-09-136-605-14
Sequence 14, Application US/09136605A
Patent No. 6140052
GENERAL INFORMATION:
APPLICANT: He, Tong-Chuan
APPLICANT: Kinzler, Kenneth
APPLICANT: Vogelstein, Bert
TITLE OF INVENTION: Beta Catenin, TCF-4, and APC Interact to
TITLE OF INVENTION: Prevent Cancer
FILE REFERENCE: 1107.75741
CURRENT APPLICATION NUMBER: US/09/136,605A
CURRENT FILING DATE: 1998-08-20
EARLIER APPLICATION NUMBER: 08/821,355
EARLIER FILING DATE: 1997-03-20
EARLIER APPLICATION NUMBER: 09/003,687
EARLIER FILING DATE: 1998-01-06
NUMBER OF SEQ ID NOS: 28
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 14
LENGTH: 8056
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: TATA_signal
LOCATION: (2458)...(2462)
US-09-136-605-14

Query Match   27.9%;   Score 26.8; DB 3; Length 8056;
Best Local Similarity   61.4%; Pred. No. 2.4;
Matches   43; Conservative   0; Mismatches   27; Indels   0; Gaps   0;

```
Qy    17  ggggcgccttggccatggagtcggaggaaataagaaagaggacgattggctgtcgatggtgct  76
          ||||| |||||||  |||||| |||||| ||||||||||||||||| ||||||  ||||||||||
Db  5570  ggggcggggtggcaggagtcgtatgaatgaggattaagagaggattgatctctgagagtga  5629

Qy    77  cagaactgct  86
          ||| |||||
Db  5630  atgaattgct  5639
```

RESULT 3
US-08-306-691B-4i
Sequence 41, Application US/08306691B
Patent No. 5734039
GENERAL INFORMATION:
APPLICANT: Calabretta, Bruno
APPLICANT: Skorski, Tomasz
TITLE OF INVENTION: ANTISENSE
TITLE OF INVENTION: OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seidel, Gonda, Lavorgna & Monaco, P.C.
STREET: Two Penn Center, Suite 1800
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: U.S.A.
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/306,691B
FILING DATE: September 15, 1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Monaco, Daniel A.
REGISTRATION NUMBER: 30,480
REFERENCE/DOCKET NUMBER: 8321-8
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-8383
TELEFAX: (215) 568-5549
TELEX: No. 5734039e
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 8082 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-306-691B-41

Query Match   27.9%;   Score 26.8; DB 1; Length 8082;
Best Local Similarity   61.4%; Pred. No. 2.4;
Matches   43; Conservative   0; Mismatches   27; Indels   0; Gaps   0;

```
Qy    17  ggggcgccttggccatggagtcggaggaaataagaaagaggacgattggctgtcgatggtgct  76
          ||||| |||||||  |||||| |||||| ||||||||||||||||| ||||||  ||||||||||
Db  5596  GGGGCGGGGTGGCAGGGAGTCGTPATGAATGAGGATTAAGAGAGGATTGATCTCTCGAGAGTGA  5655

Qy    77  cagaactgct  86
          ||| |||||
Db  5656  ATGAATTGCT  5665
```

RESULT 4
US-08-187-785-1
Sequence 1, Application US/08187785
Patent No. 5756476
GENERAL INFORMATION:
APPLICANT: Epstein, Stephen
APPLICANT: Unger, Ellis
APPLICANT: Speir, Edith
TITLE OF INVENTION: Inhibition of No. 5756476-Transformed Cell
TITLE OF INVENTION: Proliferation Using Anti-Sense Oligonucleotides
NUMBER OF SEQUENCES: 13
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/187,785
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/821,415
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Altman, Daniel E.
REGISTRATION NUMBER: 34,115
REFERENCE/DOCKET NUMBER: NIH001.001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 714-760-0404
TELEFAX: 714-760-9502

CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson, and Bear
STREET: 620 Newport Center Dr. Sixteenth Floor
CITY: Newport Beach
STATE: CA
COUNTRY: USA
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 8082 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   ORIGINAL SOURCE:
;   INDIVIDUAL ISOLATE: Human
;   IMMEDIATE SOURCE:
;   CLONE: C-Myc Genomic Clone
US-08-187-785-1

Query Match   Best Local Similarity   27.9%; Score 26.8; DB 1; Length 8082;
Matches   43; Conservative   0; Mismatches   27; Indels   0; Gaps   0;

```
Qy    17 ggggcgccttggcatggagtggaggaataagaaaaggaggacggattggcttgtcgatggtgct 76
         ||||||  ||||||| |||||||||||||||  ||||||| |||||| || ||||| ||  |||
Db  5596 GGGGCGGGGTGGCAGGGAGTGTATGAATGAGGATAAGAGAGGAGGATTGATCTCTGAGAGTGA 5655

Qy    77 cagaactgct 86
         ||| ||||
Db  5656 ATGAATTGCT 5665
```

RESULT 5
PCT-US93-06251-28
; Sequence 28, Application PC/TUS9306251
; GENERAL INFORMATION:
;   APPLICANT: Wickstrom, Eric and Rife, Jason P.
;   TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing
;   TITLE OF INVENTION: stereospecific Alkylphosphonates and Arylphosphonates
;   NUMBER OF SEQUENCES: 93
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
;   STREET: 400 Garden City Plaza
;   CITY: Garden City
;   STATE: NY
;   COUNTRY: USA
;   ZIP: 11530
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US93/06251
;   FILING DATE: 19930630
;   CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: DiGiglio, Frank S.
;   REGISTRATION NUMBER: 31,346
;   REFERENCE/DOCKET NUMBER: 8586
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 516-742-4343
;   TELEFAX: 516-742-4366
;   TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 28:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 8082 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
PCT-US93-06251-28

Query Match   Best Local Similarity   27.9%; Score 26.8; DB 5; Length 8082;

Matches   43; Conservative   0; Mismatches   27; Indels   0; Gaps   0;

```
Qy    17 ggggcgccttggcatggagtggaggaataagaaaaggaggacggattggcttgtcgatggtgct 76
         ||||||  ||||||| |||||||||||||||  ||||||| |||||| || ||||| ||  |||
Db  5596 GGGGCGGGGTGGCAGGGAGTGTATGAATGAGGATAAGAGAGGAGGATTGATCTCTGAGAGTGA 5655

Qy    77 cagaactgct 86
         ||| ||||
Db  5656 ATGAATTGCT 5665
```

RESULT 6
US-09-018-584A-43
; Sequence 43, Application US/0901858A4
; Patent No. 6238863
; GENERAL INFORMATION:
;   APPLICANT: Schumm, James W.
;   APPLICANT: Bacher, Jeffery W.
;   TITLE OF INVENTION: MATERIALS AND METHODS FOR
;   TITLE OF INVENTION: IDENTIFYING AND ANALYZING INTERMEDIATE TANDEM
;   TITLE OF INVENTION: REPEAT DNA MARKERS
;   NUMBER OF SEQUENCES: 147
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Promega Corporation
;   STREET: 2800 Woods Hollow Road
;   CITY: Madison
;   STATE: Wisconsin
;   COUNTRY: U.S.A.
;   ZIP: 53711-5399
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Mb
;   COMPUTER: IBM compatible PC
;   OPERATING SYSTEM: Windows 95
;   SOFTWARE: Word 97 (DOS text format)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/018,584A
;   FILING DATE: 04-Feb-1998
;   CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Grady J. Frenchick
;   REGISTRATION NUMBER: 29,018
;   REFERENCE/DOCKET NUMBER: 16026.9180
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (608) 257-3501
;   TELEFAX: (608) 257-2275
; INFORMATION FOR SEQ ID NO: 43:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1366 bp
;   TYPE: Nucleic Acid
;   STRANDEDNESS: Double
;   TOPOLOGY: Circular
;   MOLECULE TYPE: Genomic DNA
;   HYPOTHETICAL: no
;   IMMEDIATE SOURCE:
;   CLONE: S110
US-09-018-584A-43

Query Match   Best Local Similarity   27.7%; Score 26.6; DB 4; Length 1366;
Matches   47; Conservative   0; Mismatches   34; Indels   0; Gaps   0;

```
Qy    15 tcgggggcgccttggcatggagtggaggaataagaaaaggaggacggattggcttgtcgatggtg 74
         ||||| |||| || ||| |||||| ||| |||| |||| || ||| |||  ||||  || ||
Db   164 TCGGAGCTGCAGTGAGAAAGCGAAAGGAATGGAAGACACAGAATGGAATGGAATGGCACTGGA 223

Qy    75 ctcagaactgctggagtggag 95
         ||||| |||||||||||||||
Db   224 ATGGAATGGAGTGGAGTGGAG 244
```

RESULT 7
US-08-588-985-1

; Sequence 1, Application US/08588985
; Patent No. 5777094
; GENERAL INFORMATION:
; APPLICANT: Michiyuki MATSUDA et al.
; TITLE OF INVENTION: cDNA OF DOCK180 GENE AND DOCK180 PROTEIN
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/588,985
; FILING DATE: January 19, 1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6519 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE: spleen cell of homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 24..5619
US-08-588-985-1

Query Match   27.5%;   Score 26.4; DB 1; Length 6519;
Best Local Similarity   61.8%;   Pred. No. 3.1;
Matches   42; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy   26 tggcatggagtggaggaataagaaaagaggagcgattggctcgatggtgtgctcagaactgc 85
         ||||||||||||| |||| |||||| ||| ||||||||| ||| || ||| ||| || ||
Db   938 TCCCATGGAGCTGAGGGACAACAACACCAGGAAAACTCACCTCGGGGTTGCGCGGCGACCTTT 997

Qy   86 tggagtgg 93
         ||||||||
Db   998 TGGAGTGG 1005

RESULT   8
US-08-971-988-1
; Sequence 1, Application US/08971988
; Patent No. 5786461
; GENERAL INFORMATION:
; APPLICANT: Michiyuki MATSUDA et al.
; TITLE OF INVENTION: cDNA OF DOCK180 GENE AND DOCK180 PROTEIN
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/971,988
; FILING DATE: 17-NOV-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/588,985
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6519 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE: spleen cell of homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 24..5619
US-08-971-988-1

Query Match   27.5%;   Score 26.4; DB 1; Length 6519;
Best Local Similarity   61.8%;   Pred. No. 3.1;
Matches   42; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy   26 tggcatggagtggaggaataagaaaagaggagcgattggctcgatggtgtgctcagaactgc 85
         ||||||||||||| |||| |||||| ||| ||||||||| ||| || ||| ||| || ||
Db   938 TCCCATGGAGCTGAGGGACAACAACACCAGGAAAACTCACCTCGGGGTTGCGCGGCGACCTTT 997

Qy   86 tggagtgg 93
         ||||||||
Db   998 TGGAGTGG 1005

RESULT   9
US-08-434-000A-5/c
; Sequence 5, Application US/08434000A
; Patent No. 6046037
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071

~ COMPUTER READABLE FORM:
~ MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
~ MEDIUM TYPE: storage
~ COMPUTER: IBM Compatible
~ OPERATING SYSTEM: IBM P.C. DOS 5.0
~ SOFTWARE: Word Perfect 5.1
~ CURRENT APPLICATION DATA:
~ APPLICATION NUMBER: US/08/434,000A
~ FILING DATE:
~ CLASSIFICATION: 435
~ PRIOR APPLICATION DATA:
~ PRIOR APPLICATION DATA:
~ PRIOR APPLICATION DATA: including application
~ APPLICATION NUMBER: 08/367,395
~ FILING DATE: 12/30/94
~ ATTORNEY/AGENT INFORMATION:
~ NAME: Guise, Jeffrey W.
~ REGISTRATION NUMBER: 34,613
~ REFERENCE/DOCKET NUMBER: 212/127
~ TELECOMMUNICATION INFORMATION:
~ TELEPHONE: (619) 552-8400
~ TELEFAX: (619) 552-0159
~ TELEX: 67-3510
~ INFORMATION FOR SEQ ID NO: 5:
~ SEQUENCE LISTING
~ SEQUENCE CHARACTERISTICS:
~ LENGTH: 3630 base pairs
~ TYPE: nucleic acid
~ STRANDEDNESS: single
~ TOPOLOGY: linear
~ FEATURE:
~ NAME/KEY: Coding Sequence
~ LOCATION: 152...2425
~ DESCRIPTION: Bovine Polyimmunoglobulin Receptor
US-08-434-000A-5

Query Match   26.9%;   Score 25.8; DB 3; Length 3630;
Best Local Similarity   58.4%; Pred. No. 4.1;
Matches   45; Conservative   0; Mismatches   32; Indels   0; Gaps   0;

Qy   20 gcgccttggcatggagtggaggaataagaaaagaaggagcgcgattggctgatggtgctcag 79
          | | || | ||||| | |||||| | || | |||| | | || |||| | | | |||||| |
Db   2881 GCACATTTTAATGAGGGAAAGAAGAAGAAGAAGAAGATGGAATGGAGTGGGGATGGAGAGGTG 2822

Qy   80 aactgctggagtggagg 96
          | ||| | |||| ||||
Db   2821 CAGAGATGGGGTGGAAG 2805

RESULT 10
US-09-312-157-5/c
Sequence 5, Application US/09312157
Patent No. 6303341
GENERAL INFORMATION:
APPLICANT: ANDREW C. HIATT, JULIAN
K.-C. MA, THOMAS LEHNER
TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
PROTEINS IN PLANTS AND THEIR USES
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0

~ SOFTWARE: Word Perfect 5.1
~ CURRENT APPLICATION DATA:
~ APPLICATION NUMBER: US/09/312,157
~ FILING DATE: 14-May-1999
~ CLASSIFICATION: <Unknown>
~ PRIOR APPLICATION DATA:
~ APPLICATION NUMBER: 08/434,000
~ FILING DATE: <Unknown>
~ ATTORNEY/AGENT INFORMATION:
~ NAME: Guise, Jeffrey W.
~ REGISTRATION NUMBER: 34,613
~ REFERENCE/DOCKET NUMBER: 212/127
~ TELECOMMUNICATION INFORMATION:
~ TELEPHONE: (619) 552-8400
~ TELEFAX: (619) 552-0159
~ TELEX: 67-351
~ INFORMATION FOR SEQ ID NO: 5:
~ SEQUENCE LISTING
~ SEQUENCE CHARACTERISTICS:
~ LENGTH: 3630 base pairs
~ TYPE: nucleic acid
~ STRANDEDNESS: single
~ TOPOLOGY: linear
~ FEATURE:
~ NAME/KEY: Coding Sequence
~ LOCATION: 152...2425
~ SEQUENCE DESCRIPTION: SEQ ID NO: 5:
~ DESCRIPTION: Bovine Polyimmunoglobulin Receptor
US-09-312-157-5

Query Match   26.9%;   Score 25.8; DB 4; Length 3630;
Best Local Similarity   58.4%; Pred. No. 4.1;
Matches   45; Conservative   0; Mismatches   32; Indels   0; Gaps   0;

Qy   20 gcgccttggcatggagtggaggaataagaaaagaaggagcgcgattggctgatggtgctcag 79
          | | || | ||||| | |||||| | || | |||| | | || |||| | | | |||||| |
Db   2881 GCACATTTTAATGAGGGAAAGAAGAAGAAGAAGAAGATGGAATGGAGTGGGGATGGAGAGGTG 2822

Qy   80 aactgctggagtggagg 96
          | ||| | |||| ||||
Db   2821 CAGAGATGGGGTGGAAG 2805

RESULT 11
PCT-US93-06251-23
Sequence 23, Application PCT/US93306251
GENERAL INFORMATION:
APPLICANT: Wickstrom, Eric and Rife, Jason P.
TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing
TITLE OF INVENTION: Stereospecific Alkylphosphonates and Arylphosphonates
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/06251
FILING DATE: 19930630
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: DiGiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8586
TELECOMMUNICATION INFORMATION:

;   TELEPHONE: 516-742-4343
;   TELEFAX: 516-742-4366
;   TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 23:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 11558 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
PCT-US93-06251-23

Qy   6 cttttctgtcgggcggccttggcatcggatggagggaataagaaaagacgaatggctg 65
          ||||||| ||||||||||| ||| |||||||||||| ||| |||| || ||||||
Db   4930 CTTCTGCCATAGGGGAGCCAGGATTTGGAGAGAGACAGAGAATAAAGCATGACATGAGGGGGTA 4989

Qy   66 tcgatggtgctcagaactgctggag 90
          || || ||||| |||| ||||||
Db   4990 CCAGTGGTTCAGACCCACCTGGG 5014

Query Match   26.9%;   Score 25.8;   DB 5;   Length 11558;
Best Local Similarity 56.5%;   Pred. No. 6.2;
Matches   48;   Conservative 0; Mismatches   37; Indels   0; Gaps   0;

RESULT 12
US-08-518-878B-36/c
Sequence 36, Application US/08518878B
Patent No. 5702902
GENERAL INFORMATION:
 APPLICANT: Tartaglia, Louis A.
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
 TITLE OF INVENTION: TREATMENT OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
 NUMBER OF SEQUENCES: 57
 CORRESPONDENCE ADDRESS:
   ADDRESSEE: Pennie & Edmonds
   STREET: 1155 Avenue of the Americas
   CITY: New York
   STATE: New York
   COUNTRY: U.S.A.
   ZIP: 10036-2711
 COMPUTER READABLE FORM:
   MEDIUM TYPE: Floppy disk
   COMPUTER: IBM PC compatible
   OPERATING SYSTEM: PC-DOS/MS-DOS
   SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/08/518,878B
   FILING DATE: 23-AUG-1995
   CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
   NAME: Coruzzi, Laura A.
   REGISTRATION NUMBER: 30,742
   REFERENCE/DOCKET NUMBER: 7853-036
 TELECOMMUNICATION INFORMATION:
   TELEPHONE: (212) 790-9090
   TELEFAX: (212) 869-9741/8864
   TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 36:
   SEQUENCE CHARACTERISTICS:
     LENGTH: 1205 base pairs
     TYPE: nucleic acid
     STRANDEDNESS: single
     TOPOLOGY: linear
US-08-518-878B-36

Qy   35 gtggaggaataagaaaagagacgattggctgtcgatggtgctcagaactgctggagtgga 94

Db   1133 GGGGATGAAGATGAAGATGTGGGCTTAGCTGTAGAAAGGGCTGAGGGCTCAGGGTCTGGT 1074
          ||| ||| ||| | ||||| || ||| ||||| ||| ||||| ||| ||||| || |||
Qy   95 g 95
          |
Db   1073 G 1073

RESULT 13
US-08-294-522B-37/c
Sequence 37, Application US/08294522B
Patent No. 5741666
GENERAL INFORMATION:
 APPLICANT: Tartaglia, Louis A.
 TITLE OF INVENTION: Compositions and Methods for the
 TITLE OF INVENTION: Treatment of Body Weight Disorders, Including Obesity
 NUMBER OF SEQUENCES: 48
 CORRESPONDENCE ADDRESS:
   ADDRESSEE: Pennie & Edmonds
   STREET: 1155 Avenue of the Americas
   CITY: New York
   STATE: New York
   ZIP: 10036-2711
 COMPUTER READABLE FORM:
   MEDIUM TYPE: Floppy disk
   COMPUTER: IBM PC compatible
   OPERATING SYSTEM: PC-DOS/MS-DOS
   SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/08/294,522B
   FILING DATE: 23-AUG-1994
   CLASSIFICATION: 514
 ATTORNEY/AGENT INFORMATION:
   NAME: Coruzzi, Laura A.
   REGISTRATION NUMBER: 30,742
   REFERENCE/DOCKET NUMBER: 7853-015
 TELECOMMUNICATION INFORMATION:
   TELEPHONE: (212) 790-9090
   TELEFAX: (212) 869-8864/9741
 INFORMATION FOR SEQ ID NO: 37:
   SEQUENCE CHARACTERISTICS:
     LENGTH: 1205 base pairs
     TYPE: nucleic acid
     STRANDEDNESS: both
     TOPOLOGY: linear
   MOLECULE TYPE: cDNA
   ANTI-SENSE: NO
   FEATURE:
     NAME/KEY: CDS
     LOCATION: 1..912
US-08-294-522B-37

Query Match   25.2%;   Score 24.2;   DB 1;   Length 1205;
Best Local Similarity 62.3%;   Pred. No. 10;
Matches   38;   Conservative 0; Mismatches   23; Indels   0; Gaps   0;

Qy   35 gtggaggaataagaaaagagacgattggctgtcgatggtgctcagaactgctggagtgga 94
          ||| |||| ||| | ||||| || ||| || ||||| |||| || ||||| || ||
Db   1133 GGGGATGAAGATGAAGATGTGGGCTTAGCTGTAGAAAGGCTGAGGGCTCAGGGTCTGGT 1074
Qy   95 g 95
          |
Db   1073 G 1073

RESULT 14
US-08-807-861A-36/c
Sequence 36, Application US/08807861A
Patent No. 5853975
GENERAL INFORMATION:
 APPLICANT: Tartaglia, Louis A.

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
TITLE OF INVENTION: REGULATION OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/807,861A
FILING DATE: 26-FEB-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/518,878
FILING DATE: 23-AUG-1995
APPLICATION NUMBER: US 08/470,868
FILING DATE: 06-JUN-1995
APPLICATION NUMBER: US 08/294,522
FILING DATE: 23-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-066
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 1205 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-807-861A-36

Query Match   25.2%;   Score 24.2;   DB 2;   Length 1205;
Best Local Similarity   62.3%;   Pred. No. 10;
Matches   38; Conservative   0; Mismatches   23; Indels   0; Gaps   0;

```
Qy   35 gtggaggaataagaaaagagagcgattggcctgtcgatggtgctcagaactgctggagtgga 94
         |||| ||| || ||| || |  ||||  | | || | |   |  |  ||  ||   |||| ||
Db 1133 GGGGATGAAGATGAAGATGTGGGCTTAGCTGTGTAGAAAAGGGCTGAGGGCTCAGGGTCTGGT 1074

Qy   95 g 95
         |
Db 1073 G 1073
```

RESULT 15
US-08-470-868A-36/c
Sequence 36, Application US/08470868A
Patent No. 5861485
GENERAL INFORMATION:
APPLICANT: Tartaglia, Louis C.
TITLE OF INVENTION: Compositions and Methods for the
TITLE OF INVENTION: Treatment of Body Weight Disorders, Including Obesity
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie and Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,868A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-0031-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66441 PENNIE
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 1205 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-470-868A-36

Query Match   25.2%;   Score 24.2;   DB 2;   Length 1205;
Best Local Similarity   62.3%;   Pred. No. 10;
Matches   38; Conservative   0; Mismatches   23; Indels   0; Gaps   0;

```
Qy   35 gtggaggaataagaaaagagagcgattggcctgtcgatggtgctcagaactgctggagtgga 94
         |||| ||| || ||| || |  ||||  | | || | |   |  |  ||  ||   |||| ||
Db 1133 GGGGATGAAGATGAAGATGTGGGCTTAGCTGTGTAGAAAAGGGCTGAGGGCTCAGGGTCTGGT 1074

Qy   95 g 95
         |
Db 1073 G 1073
```

Search completed: February 22, 2002, 08:29:04
Job time: 3114 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 22, 2002, 07:37:10 ; Search time 1531.74 Seconds
               (without alignments)
               673.479 Million cell updates/sec

Title: US-09-509-482-1
Perfect score: 96
Sequence: 1 gtttccttcctctgtcgggg............cagaactgctgggagtggagg 96

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues
               Listing first 45 summaries
               Minimum Match 100%
               Maximum Match 100%

Post-processing: Minimum Match 0%

Maximum DB seq length: 2000000000
Minimum DB seq length: 0

Total number of hits satisfying chosen parameters: 22703874

Database :   EST:*
               1: em_estfun:*
               2: em_esthum:*
               3: em_estin:*
               4: em_estom:*
               5: em_estpl:*
               6: em_estba:*
               7: em_estro:*
               8: em_estov:*
               9: em_htc:*
               10: em_htc:*
               11: gb_esti:*
               11: gb_est2:*
               12: gb_htc:*
               13: gb_gss:*
               14: em_gss_fun:*
               15: em_gss_hum:*
               16: em_gss_inv:*
               17: em_gss_pln:*
               18: em_gss_pro:*
               19: em_gss_rod:*
               20: em_gss_vrt:*
               21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	29.8	31.0	313	13	BH019043	L2252b.d_
c 2	29.4	30.6	764	13	CNS01OAR	Anopheles
c 3	29.4	30.6	764	11	BF599968	60205267B
c 4	29	30.2	366	11	BF249184	NcEST3a08
c 5	28.8	30.0	219	10	AA914519	vy95f01.r
c 6	28.8	30.0	432	13	AA914519	vy95f01.r
c 7	28.6	29.8	766	13	B04814	B04814 CSRL-46H1-u
c 8	28.6	29.6	456	10	AO751154	HS_5575_B
c 9	28.4	29.6	512	10	AA161741	MBAFCF1H1
c 10	28.4	29.6	706	11	AA445901	SWAMCA185
c 11	28.4	29.6	801	11	BG269160	L0-317GT3
c 12	28.2	29.4	849	13	CNS01WH4	AL170401 Tetraodon
c 13	28	29.2	296	10	BB021970	BB021970
c 14	28	29.2	498	13	AO881871	HS_5304_A
c 15	28	29.2	666	13	AZ349663	1M00865I4
c 16	28	29.2	845	11	CNS05HYF	AL338208 Tetraodon
c 17	27.8	29.0	221	10	AA886991	oj29g12.s
c 18	27.8	29.0	434	10	AW587661	ST763B05 P
c 19	27.8	29.0	457	11	BG413008	1a84a03.y
c 20	27.6	28.8	470	11	BG173725	602336868
c 21	27.6	28.8	569	11	BG833218	951002D06
c 22	27.6	28.8	571	11	BG833227	951002D12
c 23	27.6	28.8	744	13	BH079141	RPCI-24-2
c 24	27.6	28.8	864	11	BG442094	GA__Ea001
c 25	27.6	28.8	1101	11	AL068664	Drosophil
c 26	27.4	28.5	129	10	AV165065	AV165065
c 27	27.4	28.5	289	13	AQ248809	T15K16-SP
c 28	27.4	28.5	311	10	BB318465	BB318465
c 29	27.4	28.5	476	13	AQ121717	HS_3087_B
c 30	27.4	28.5	501	13	AQ154930	HS_3038_A
c 31	27.4	28.5	804	13	AZ176215	SP_0141_A
c 32	27.2	28.3	354	10	AA324941	EST27851
c 33	27.2	28.3	361	11	T82914	yd39d06.r1
c 34	27.2	28.3	379	10	AA434135	zw28a07.s
c 35	27.2	28.3	395	10	AA045416	zk559b06.r
c 36	27.2	28.3	409	10	AA448807	zx10g06.r
c 37	27.2	28.3	436	10	BE128749	DEPA2499
c 38	27.2	28.3	444	10	AW226368	ST81E08 P
c 39	27.2	28.3	464	10	BE278496	601156883
c 40	27.2	28.3	526	10	BE265917	601191184
c 41	27.2	28.3	532	11	BG403683	602419510
c 42	27.2	28.3	579	11	BF836271	QV3-HT101
c 43	27.2	28.3	618	11	BE615867	601279740
c 44	27.2	28.3	715	11	BG674716	602620820
c 45	27.2	28.3	716	11	BF568539	602184155

ALIGNMENTS

RESULT 1
BH019043/c
BH019043/c
BH019043 L2252b.d_HygT7a.2 Leishmania major Friedlin Cosmid Genomic Genomic Library

LOCUS   BH019043   313 bp   DNA   GSS   25-MAY-2001
DEFINITION L2252b.d.HygT7a.2 Leishmania major genomic clone L2252b, DNA sequence.
ACCESSION   BH019043
VERSION   BH019043.1 GI:14197749
KEYWORDS   GSS.
SOURCE   Leishmania major
ORGANISM   Leishmania major
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1 (bases 1 to 313)
AUTHORS   Myler,P.J., Vogt,C., Cawthra,J., Klacking,M., Marty,A., Mack,J.,
            Munden,H., Nguyen,D., Robertson,L., Sisk,E., Fazelinia,G., Aggarwal
            ,G., Nelson,S., Seyler,A., Worthey,E. and Stuart,K.
TITLE   Leishmania major Friedlin Cosmid End Sequences
JOURNAL   Unpublished (2000)
COMMENT   Other_GSSs: L2252b.d_HygT7a.1
            Contact: Myler PJ
            Seattle Biomedical Research Institute
            4 Nickerson Street, Seattle, WA 98109-1651, USA
            Tel: 206 284-8846
            Fax: 206 284-0313
            Email: mylerp@sbri.org
            Seq primer: HygT7a
            Class: cosmid ends.
FEATURES   Location/Qualifiers
     source   1..313
                     /organism="Leishmania major"
                     /strain="Friedlin"
                     /db_xref="taxon:5664"
                     /clone="L2252b"
                     /clone_lib="Leishmania major Friedlin Cosmid Genomic
                     Library"

BASE COUNT   62 a   97 c   90 g   64 t
ORIGIN

/lab_host="E. coli ED8767"
/note="Vector; cLHYG; Site_1; BamHI; Genomic DNA from Leishmania major Friedlin was partially digested with Sau3AI, size selected, and ligated with BamHI-digested cLHYG cosmid vector DNA. 9216 clones were picked and arrayed. Library construction is described in Ivens et al, Genomics Research, 8:135-145 (1998). The cLHYG vector (Acc. No. CVU59231) is described in Ryan et al, Gene, 131:145-150 (1993)."

Qy   35 gtgaggaaataagagaagcagcgattggctcgtcgatggtgctccagaactcgtcggagt 91
         | ||| ||||| | ||||| | ||||| ||||| ||||||||||||||||||||||||||
Db 168 GCGCCGGAGCTACGAAAAGGAGCGCGGTCGTTGCTTGTCGTTGGTCGCCCCGCCATTGCGCGCAAT 112

RESULT 2

CNS01OAR
LOCUS   CNS01OAR   764 bp   DNA   GSS   14-JUN-2001
DEFINITION Anopheles gambiae GSS T7 end of clone 25015 of NotreDamel library from strain PEST of Anopheles gambiae (African malaria mosquito), genomic survey sequence.
ACCESSION   AL153283
VERSION   AL153283.1 GI:7014202
KEYWORDS   GSS.
SOURCE   African malaria mosquito.
ORGANISM   Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
REFERENCE   1 (bases 1 to 764)
AUTHORS   Genoscope.
TITLE   Direct Submission
JOURNAL   Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage: BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
REFERENCE   2 (bases 1 to 764)
AUTHORS   Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
TITLE   Direct Submission
JOURNAL   Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr. Roux, Paris 75015, France
COMMENT   This clone is from an A. gambiae BAC library provided by F.H. Collins and sequenced by Genoscope in collaboration with the Laboratory of Biochem. and Biol. Molec. of Insects, Institut Pasteur.
FEATURES   Location/Qualifiers
     source   1..764
                     /organism="Anopheles gambiae"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="25015"
                     /clone_lib="NotreDamel"
                     /note="end : T7"
BASE COUNT   237 a   209 c   146 g   168 t   4 others
ORIGIN

Query Match   30.6%; Score 29.4; DB 13; Length 764;
Best Local Similarity 63.4%; Pred. No. 66;
Matches 45; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy   6 cttcttctgtcgggcgcgcttggcatggtaggaggaataagaaaggagcggattggctg 65
         ||||| || ||| |||| || |||||||||| ||||||| |||| ||||||||||||||
Db 469 CCTGGTTCTGTCTACCGGTCCGGTTCCCCATGGAAGTTCTGGAAAAATAAATTACCCGGATCCGTCC 528

Qy   66 tcgatggtgct 76
         |||||||||||
Db 529 GTGATGGTGCT 539

RESULT 3

BF539968/c
LOCUS   BF539968   956 bp   mRNA   EST   11-DEC-2000
DEFINITION 602052678F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4191907 5', mRNA sequence.
ACCESSION   BF539968
VERSION   BF539968.1 GI:11627336
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 956)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov/
            Plate: LLAM9521 row: 1 column: 20
            High quality sequence stop: 723.
FEATURES   Location/Qualifiers
     source   1..956
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4191907"
                     /clone_lib="NCI_CGAP_SG2"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1: Sal1; Site_2: Sal1; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.3 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT   230 a   244 c   212 g   270 t
ORIGIN

Query Match   30.6%; Score 29.4; DB 11; Length 956;
Best Local Similarity 58.6%; Pred. No. 67;
Matches 51; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy   5 cctcttctgtcgggcgcgccttggcatggtaggaggaataagaaaggagcggattggctg 64
         || ||| ||| | | || || |||| |||| || || || ||||||| |||| ||||
Db 920 CCCCCCTTTTTTCCCAGAATATATGGGGGGGGGGGACACGGGGGACNCCAAAAGGAAAAAAGGAAAAAATTGGGC 861

Qy   65 gtcgatggtgctccagaactcgtcggagt 91
         ||| ||||||||||||||| |||||||
Db 860 GTCTTTGGTGTCTCCAGAAAAAGGGGAAT 834

RESULT 4

BF249184
LOCUS   BF249184   366 bp   mRNA   EST   15-NOV-2000
DEFINITION NcEST3a08all.y1 Nc 1314 Tachyzoite cDNA Neospora caninum cDNA 5', mRNA sequence.
ACCESSION   BF249184
VERSION   BF249184.1 GI:11179367
KEYWORDS   EST.
SOURCE   Neospora caninum.
ORGANISM   Neospora caninum
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Sarcocystidae; Neospora.
REFERENCE   1 (bases 1 to 366)
AUTHORS   Cole,R., Fogarty,S., Tang,K., Howe,D.K., Sibley,L.D., Clifton,S.,

Left panel

AUTHORS   Marra,M., Hillier,L., Pape,D., Martin,J., Wylie,T., Theising,B.,
           Bowers,Y., Gibbons,M., Ritter,E., Bennet,J., Ronko,I.,
           Tsagareishvili,R., Fedele,M., Belaygorod,L., Franklin,C., Carr,L.M.
           , Grow,A., Maguire,L., Wadkins,J., Richey,J., Waterston,R. and
           Wilson,R.
TITLE   USDA-Washu Neospora EST Project
JOURNAL   Unpublished (2000)
COMMENT   Contact: Sandy Clifton, Ph.D. - Neospora
           USDA-WashU Neospora EST Project
           Washington University School of Medicine
           4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
           Tel: 314 286 1800
           Fax: 314 286 1810
           Email: est@watson.wustl.edu
           Contact David Sibley (toxoest@borcim.wustl.edu) for further
           information relating to organism, libraries, or clone availability.
           Seq primer: -40RP from Gibco
           High quality sequence stop: 356.
FEATURES   Location/Qualifiers
     source   1..366
                     /organism="Neospora caninum"
                     /strain="Nc-1"
                     /db_xref="taxon:29176"
                     /clone_lib="Nc 1314 Tachyzoite cDNA"
                     /dev_stage="Tachyzoite"
                     /lab_host="DH10B (GeneHog, Research Genetics, Inc.)"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI. This library was constructed by Steve Fogarty,
                     Robert Cole, and Keliang Tang at Washington University.
                     cDNAs were synthesized from poly(A)+ RNA by oligo d(T)
                     priming, size-selected and directionally cloned into the
                     Uni-ZAP XR lambda vector (Stratagene). The primary library
                     was mass excised as phagemids and rescued in SOLR cells.
                     Xhoi. The plasmid library was recovered from the SOLR cells and
                     transformed in mass into DH10B (GeneHog, Research Genetics
                     , Inc.) for sequencing. WARNING:the library may contain a
                     small percentage of contaminents from human fibroblast
                     cells."
BASE COUNT   65 a   80 c   111 g   110 t
ORIGIN

RESULT 5
AA914519/c
LOCUS   AA914519   219 bp   mRNA   EST   14-APR-1998
DEFINITION vy95f01.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
            IMAGE:1313977 5' similar to SW:CYA4_RAT P26770 ADENYLATE CYCLASE,
            TYPE IV ;, mRNA sequence.
ACCESSION   AA914519
VERSION   AA914519.1 GI:3053911
KEYWORDS   EST.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 219)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Query Match
Best Local Similarity   57.0%;
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
Score 29; DB 11; Length 366;

Qy   2 tttccttcctctgtcggggcgccttggcatggagtggaggaataagaaaagagcggattg 61
         ||| |||| || ||| || ||||| |||| ||||||| |||| |||||||| | |
Db 226 TTTCCTTGTTGTTCTTTCGGCAATGCTTCGTCCTTGTCCTTGATAAGAGAACCTCAGGCCTGCAAGG 285

Qy   62 gctgtcggatggtgctcaagaactgctggagtgga 94
         ||| || ||| |||| || |||| || |||||
Db 286 GATGGCCAATGTAGCCTTTAATTGTGGGAATGGA 318

Right panel

           Waterston,R.
TITLE   The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   Contact: Marra M/Mouse EST Project
           WashU-HHMI Mouse EST Project
           Washington University School of MedicineP
           4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
           Tel: 314 286 1800
           Fax: 314 286 1810
           Email: mouseest@watson.wustl.edu
           This clone is available royalty-free through LLNL ; contact the
           IMAGE Consortium (info@image.llnl.gov) for further information.
           MGI:684273
           Seq primer: -28m13 rev2 ET from Amersham
           High quality sequence stop: 204.
FEATURES   Location/Qualifiers
     source   1..219
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1313977"
                     /clone_lib="Soares_mammary_gland_NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
) with a modified polylinker; Site_1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGACTGCGAGCGGCCGCGAAATGGTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT   39 a   76 c   43 g   61 t
ORIGIN

RESULT 6
B04814
LOCUS   B04814   432 bp   DNA   GSS   13-JUL-1996
DEFINITION CSRL-46H1-u CSRL flow sorted Chromosome 11 specific cosmid Homo
            sapiens genomic clone CSRL-46H1, DNA sequence.
ACCESSION   B04814
VERSION   B04814.1 GI:1414092
KEYWORDS   GSS.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 432)
  AUTHORS   Evans,G.A., Burbee,D., Davies,C., Hahner,L., Oliver,T., Gilbert,M.,
            Jones,D., Ward,T., Gillilan,E., Schagemann,J., Probst,S., Harris
            ,J., DeFord,J., McFarland,J., Burzinski,K., Khan,M., Kupfer,K. and
            Garner,H.R.
TITLE   Genomic Sequence Sampled Map of Chromosome 11
JOURNAL   Unpublished (1996)

Query Match
Best Local Similarity   30.0%;
Matches 42; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
Score 28.8; Pred. No. 92; DB 10; Length 219;

Qy   30 atggagtggaggaataagaaaagagcggattggcttgtcgatggtgtgtgtcaagactgtgga 89
         ||| | || |||||||||| || ||||| || || || |||||| |||| ||| | | ||
Db 123 AAGAGGAGGAGAAAGAGAGAGAGAGAAGGAAGGTGAATGCTGTGAGGTGATGATCAGAGAGCTGGGGGGC 64

Qy   90 gtgg 93
         |||
Db   63 CTGG 60

```
COMMENT     Contact: Evans GA, Shane Probst
            McDermott Center for Human Growth and Development
            University of Texas Southwestern Medical Center At Dallas
            5323 Harry Hines Blvd, Dallas TX 75235-8591
            Tel: 214-648-1600
            Fax: 214-648-1666
            Email: gevans@utsw.swmed.edu, shane@mcdermott.swmed.edu
            PCR Primers
            FORWARD: AACAGGCACAAAAGTCTTAG
            BACKWARD: CTTACTTCTAGCACTGCCTG
            Seq primer: T7
            Class: cosmid ends
            High quality sequence stop: 432.
FEATURES             Location/Qualifiers
     source          1..432
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="cSRL-46H1"
                     /clone_lib="cSRL flow sorted Chromosome 11 specific
                     cosmid"
                     /sex="female"
                     /cell_type="chimeric hamster somatic cell hybrid"
                     /note="Vector: sCos-1; Human Chromosome 11 Specific cosmid
                     library prepared from flow sorted human Chromosome 11
                     derived from Chinese Hampster Ovary (CHO) monochromosomal
                     somatic cell hybrid, J1"
BASE COUNT       93 a    108 c    117 g     93 t     21 others
ORIGIN

Query Match
Best Local Similarity  30.0%;   Score 28.8;  DB 13;  Length 432;
Matches 48;  Conservative  0;  Mismatches 97;  Pred. No. 97;  Indels  0;  Gaps  0;

Qy   17 gggggcgccttggcatggagtggaggaataagaaaaggagcgaattggctcgatggtgct  76
        | |  | || ||||  ||||||| |||||| | || ||| || |||| |||||||||| |||
Db  228 GTGCAGTCCCCTTACGGGGGTAGAGGAATCAGGAAGAAAGGAGGAAATAGGTCCAGGTGGGGCT  287

Qy   77 cagaactgctggagtggagg  96
        ||| | | ||||||||||||
Db  288 TGGTGCTGGACGATGGAGG  307

RESULT  7
AQ751154
LOCUS       AQ751154                 766 bp    DNA             GSS       19-JUL-1999
DEFINITION  HS_5575_B2_E10_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1151 Col=20 Row=J, DNA sequence.
ACCESSION   AQ751154
VERSION     AQ751154.1  GI:5538312
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 766)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
```

```
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.hgsc.washington.edu
            Plate: 1151  row: J   column: 20
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 766.
FEATURES             Location/Qualifiers
     source          1..766
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="plate-1151 Col=20 Row=J"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      200 a    152 c    186 g    227 t      1 others
ORIGIN

Query Match
Best Local Similarity  29.8%;   Score 28.6;  DB 13;  Length 766;
Matches 43;  Conservative  0;  Mismatches 24;  Pred. No. 1.2e+02;  Indels  0;  Gaps  0;

Qy   16 cgggcgccttggcatggagtggaggaataagaaaaggagcgaattggctcgatggtgc  75
        | || |||| |||  |||  | ||||||||||||| || ||| |||| ||||| |||| |||
Db  470 CTGCCGCCACCTTTGTAAGGAATCAGGAAAAGGGAAAAATCAGCAACCTGCTGCTCCCTGGGTGC  529

Qy   76 tcagaac  82
        | | |||
Db  530 TGATCAC  536

RESULT  8
AA161741/c
LOCUS       AA161741                 456 bp    mRNA            EST       17-DEC-1996
DEFINITION  MBAFCF1H10T3 Brugia malayi adult female cDNA (SAW96MLW-BmAF) Brugia
            malayi cDNA clone AFCF1H10 5', mRNA sequence.
ACCESSION   AA161741
VERSION     AA161741.1  GI:1737625
KEYWORDS    EST.
SOURCE      Brugia malayi.
  ORGANISM  Brugia malayi
            Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
            Onchocercidae; Brugia.
REFERENCE   1  (bases 1 to 456)
  AUTHORS   Blaxter,M.L., Waterfall,M., Daub,J., Lizotte,M., Baron,L. and Jones
            ,S.J.
  TITLE     Genes expressed in adult female Brugia malayi
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Blaxter ML
            Institute of Cell, Animal and Population Biology
            University of Edinburgh
            Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
            3JT, UK.
            Tel: +44 131 650 6760
            Fax: +44 131 650 5450
            Email: mark.blaxter@ed.ac.uk
            The ABI trace of this sequence can be viewed at
            http://www.sanger.ac.uk/brugia/AFC/MBAFCF1H10T3.html
            Seq primer: T3
FEATURES             Location/Qualifiers
     source          1..456
                     /organism="Brugia malayi"
                     /db_xref="taxon:6279"
                     /clone="AFCF1H10"
                     /clone_lib="Brugia malayi adult female cDNA (SAW96MLW-BmAF
                     )"
                     /sex="female"
                     /dev_stage="adult"
```

RESULT 9
LOCUS   AA445901/c
DEFINITION   SWAMCA1856SK Brugia malayi adult male cDNA clone SWAMCA1856 5', mRNA sequence.
ACCESSION   AA445901
VERSION   AA445901.1 GI:2158566
KEYWORDS   EST.
SOURCE   Brugia malayi.
ORGANISM   Brugia malayi
               Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
               Onchocercidae; Brugia.
REFERENCE   1 (bases 1 to 512)
AUTHORS   Williams,S.A.
TITLE   Genes expressed in adult males of Brugia malayi
JOURNAL   Unpublished (1995)
COMMENT   Contact: Steven A. Williams
               Molecular Parasitology
               Smith College Department of Biological Sciences
               Department of Biological Sciences, Clark Science Center, Smith
               College, Northampton, MA, 01063, USA
               Tel: 4135853826
               Fax: 4135853786
               Email: genome@smith.edu
               Seq primer: pBluescript SK.

FEATURES   Location/Qualifiers
     source   1..512
                     /organism="Brugia malayi"
                     /strain="TRS Labs"
                     /db_xref="taxon:6279"
                     /clone_lib="Brugia malayi adult male cDNA (SAW94NL-BmAM)"
                     /lab_host="Xl1-Blue MRF'"
                     /note="Vector: lambda Unizap XR; Site_1: EcoR I; Site_2:
                     Xho I; Lymphatic filarial nematode parasite of humans.
                     mRNA was prepared from adult males of Brugia malayi
                     isolated from jirds and converted to double stranded cDNA
                     using reverse transcriptase and oligo(dT) followed by
                     RNase H and DNApol I. The library had 4.6 x 1OE6
                     independent recombinants and average insert size was 800
                     base pairs. The library was constructed by Noelle Ling.
                     The library is available from Dr. S.A. Williams, email
                     genome@smith.edu."

BASE COUNT   132 a   90 c   76 g   214 t
ORIGIN

Query Match   29.6%; Score 28.4; DB 10; Length 512;
Best Local Similarity   90.8%; Pred. No. 1.3e+02;
Matches 38; Conservative   0; Mismatches 16; Indels   0; Gaps   0;

RESULT 10
LOCUS   BG269160/c
ACCESSION   BG269160
VERSION   BG269160.1 GI:12974924
KEYWORDS   EST.
SOURCE   common ice plant.
ORGANISM   Mesembryanthemum crystallinum
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
               Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE   1 (bases 1 to 706)
AUTHORS   Cushman, J.C.
TITLE   An expressed sequence tag database for the common ice plant,
               Mesembryanthemum crystallinum
JOURNAL   Unpublished (1997)
COMMENT   Contact: Cushman JC
               Department of Biochemistry
               University of Nevada
               MS200, Reno, NV 89557-0014, USA
               Tel: 775-784-1918
               Fax: 775-784-1650
               Email: jcushman@unr.edu
               PCR Primers
               FORWARD: T7
               BACKWARD: T3
               Plate: L0-32 row: G column: 4
               Seq primer: T3
               High quality sequence stop: 350
               POLYA=NO.

FEATURES   Location/Qualifiers
     source   1..706
                     /organism="Mesembryanthemum crystallinum"
                     /db_xref="taxon:3544"
                     /clone="L0-3176"
                     /clone_lib="Ice plant Lambda Uni-Zap XR expression library
                     , 0 hours NaCl treatment"
                     /tissue_type="leaf"
                     /dev_stage="six week old"
                     /note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"

BASE COUNT   210 a   178 c   115 g   203 t
ORIGIN

Query Match   29.6%; Score 28.4; DB 11; Length 706;
Best Local Similarity   62.9%; Pred. No. 1.3e+02;
Matches 44; Conservative   0; Mismatches 26; Indels   0; Gaps   0;

RESULT 11
LOCUS   BG462076/c
DEFINITION   RST45079 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.

```
ACCESSION   BG462076
VERSION     BG462076.1  GI:13750582
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 801)
  AUTHORS   Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
            Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J.,
            Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
            Veloso,N., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J.
            and Ducar,M.
  TITLE     Creation of Genome-wide Protein Expression Libraries using Random
            Activation of Gene Expression
  JOURNAL   Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT     Contact: Scott J. Cain
            Athersys, Inc.
            3201 Carnegie Ave, Cleveland, OH 44115, USA
            Tel: 216 431 9900
            Fax: 216 361 9596
            Email: scain@athersys.com.
FEATURES             Location/Qualifiers
     source          1..801
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Athersys RAGE Library"
                     /clone="HT1080"
                     /cell_line="HT1080"
                     /note="See 'Creation of Genome-wide Protein Expression
                     Libraries using Random Activation of Gene Expression', in
                     press. Note that even though the cell type indicated is
                     HT1080, since a random activation method was used, these
                     sequence tags are not necessarily expressed in HT1080
                     under normal circumstances."
BASE COUNT      180 a    339 c    105 g    176 t    1 others
ORIGIN
```

```
RESULT 12
Query Match            29.6%;  Score 28.4;  DB 11;  Length 801;
Best Local Similarity 66.1%;   Pred. No. 1.3e+02;
Matches 41;  Conservative 0;  Mismatches 21;  Indels 0;  Gaps 0;

Qy  27 ggcatggagtggaggaataaagaaaggaggacgagtgtcgatgggtgctcagaaactgct 86
          |||  ||| ||| || ||| ||| |||  |||| || | |  | ||| ||| || |||
Db 767 GGGCGTGGTGGTGGGTGGGTGGAGTAGGAAAAGAGGGCGATTGATGGCAATGGCAGGGGGGAACTATT 708

Qy  87 gg 88
       ||
Db 707 GG 706
```

```
ACCESSION   AL170401
VERSION     AL170401.1  GI:7808458
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 849)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
LOCUS       CNS01WH4         849 bp    DNA        GSS       12-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            200B19 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
```

```
REFERENCE   2  (bases 1 to 849)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 849)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
            This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1..849
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="200B19"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : COAG200CA10LP1-end : T7"
BASE COUNT       68 c    314 g    135 t    45 others
ORIGIN
```

```
RESULT 13
Query Match            29.4%;  Score 28.2;  DB 13;  Length 849;
Best Local Similarity 59.2%;   Pred. No. 1.5e+02;
Matches 42;  Conservative 3;  Mismatches 26;  Indels 0;  Gaps 0;

Qy  26 tggcatggagtggaggaataaagaaaggaggacgagtgtcgatgggtggtctcagaaactgc 85
         | |  ||| ||| || |||  ||  || ||| || |||  |||  || ||| ||  || ||
Db 503 TGAGGTGGTGAGATGAGAGAVGAGAVGAGAWATAGAGGTGATTKGGTGATGAGGGAGGTGAGAWGAGT 562

Qy  86 tggagtggagg 96
       ||  |||||||
Db 563 TGAGGTGGAGG 573
```

```
ACCESSION   BB021970
VERSION     BB021970.1  GI:8194383
KEYWORDS    EST.
SOURCE      Mus musculus
  ORGANISM  Mus musculus
            house mouse.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 296)
  AUTHORS   Konno,H., Aizawa,K., Akiyama,S., Akiyama,J., Arakawa,T., Carninci,
            P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,Y.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,
            Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.,
            Suzuki,H., Suzuki,Y., Tagawa,A., Takahashi,F., Tominaga,N., Toya,
            T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamanaka,I., Yamanaka,I.,
            Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,
            M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
LOCUS       BB021970          296 bp    mRNA       EST       23-JUN-2000
DEFINITION  BB021970 RIKEN full-length enriched, adult male pituitary gland Mus
            musculus cDNA clone 5330408F08 3', mRNA sequence.
```

```
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
            Email: genome-res@gsc.riken.go.jp;
            URL:http://genome.gsc.riken.go.jp/
FEATURES             Location/Qualifiers
     source          1..296
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="5330408F08"
                     /clone_lib="RIKEN full-length enriched, adult male
                     pituitary gland"
                     /sex="male"
                     /tissue_type="pituitary gland"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'-
                     GAGAGAGAGAGAGAGACTCTTTTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 3.0 and subtraction to Rot = 100.0. Second strand
                     cDNA was prepared with the primer adapter of sequence [5',
                     GAGAGAGAGAGATTCTCGAGTTAATTAAATTAANTCCCCCCCCCCCCCC 3'. cDNA
                     was cloned into the XhoI and BamHI sites. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."
BASE COUNT       55 a    121 c     38 g     82 t
ORIGIN
```

```
RESULT 14
AQ081871

Query Match            29.2%;  Score 28;  DB 10;  Length 296;
Best Local Similarity  60.5%;  Pred. No. 1.6e+02;
Matches  46;  Conservative  0;  Mismatches 30;  Indels  0;  Gaps  0;

Qy    20  gcgccttgagctgagtggaggaataaagggagccgtcgcatggtgtctcag   79
          ||   | |||||||||||||||||||||| |||||
Db   257  GGGCTGGGGTGGGCTGGGGGATGAAAAAGAAAAAGTGGGCTTAGCTGTAGAAAGGGCTGAG  198

Qy    80  aactgctgaagtggag   95
          ||  ||| ||||
Db   197  GGCTCAGGGTCTGTG  182

LOCUS       AQ081871      498 bp      DNA          GSS       09-NOV-1999
DEFINITION  HS_5304_A1_E05_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=9072 Col=9 Row=I, DNA sequence.
ACCESSION   AQ081871
VERSION     AQ081871.1  GI:6313338
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 498)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 9072  row: I  column: 9
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 498.
FEATURES             Location/Qualifiers
     source          1..498
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=9072 Col=9 Row=I"
                     /sex="male"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT  120 a   138 c   112 g   124 t    4 others
ORIGIN
```

```
RESULT 15
AZ349663

Query Match            29.2%;  Score 28;  DB 13;  Length 498;
Best Local Similarity  55.38%;  Pred. No. 1.7e+02;
Matches  52;  Conservative  0;  Mismatches 42;  Indels  0;  Gaps  0;

Qy    1  gtttcctttctctgtcgggcggccccttggcatggagtggaggaataaagggagccgatt   60
         ||||| |||| |||| |  | |||| |||  ||||||||||||||||||| ||
Db  323  GTCTCAAGCTTCAGGATGGCGCCCTCGCANGGACTCATGGNATGAGATGTCCAANGGCC  382

Qy   61  ggctgtcgatggtgctccagaactgctggagtgga   94
         |||||| ||| || ||| ||||||| ||| || ||
Db  383  TCTTCTCACTCGTCCTCTACTCTACCAGAGGGCGCA  416

LOCUS       AZ349663      666 bp      DNA          GSS       29-SEP-2000
DEFINITION  1M0086E14R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0086E14 R, DNA sequence.
ACCESSION   AZ349663
VERSION     AZ349663.1  GI:10428900
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 666)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
```

TITLE   Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL   Unpublished (2000)
COMMENT   Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0086   row: E column: 14
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 666.
FEATURES   Location/Qualifiers
     source   1..666
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0086E14"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain Xl10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male), was obtained from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1|), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli Xl10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT   163 a   163 c   161 g   179 t
ORIGIN

Query Match   29.2%;   Score 28;   DB 13;   Length 666;
Best Local Similarity   58.3%;   Pred. No. 1.7e+02;
Matches   49;   Conservative   0;   Mismatches   35;   Indels   0;   Gaps   0;

Qy   9 ctctctgtcggggcgcctggcatggagtggaggaataagaaaggagcggattggctgtcg 68
           |||||| ||| || |||||| | |||||| |||| |||| || | || ||||
Db   358 CTTCTGCTCGGCAGGGAGGCAAACAGATGGGGGTTCAGAAGTTCAGGAATGGTGTCTCGC 417

Qy   69 atggtgctcagaactgctggagtg 92
           | | |||| || ||||| ||| ||
Db   418 AAGGTACTCAGAAGTGCAGGACTG 441